

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	258	92.5	853	3	US-09-254-325-2	Sequence 2, Appli
2	258	92.5	866	3	US-09-079-415-6	Sequence 6, Appli
3	258	92.5	866	3	US-08-750-458A-2	Sequence 2, Appli
4	188	67.4	449	2	US-08-927-394-2	Sequence 2, Appli
5	54	19.4	108	1	US-08-204-740-11	Sequence 11, Appli
6	54	19.4	108	3	US-09-081-167A-11	Sequence 11, Appli
7	54	19.4	108	3	US-09-081-395-11	Sequence 11, Appli
8	54	19.4	108	4	US-09-416-833-11	Sequence 11, Appli
9	54	19.4	108	5	PCT-US95-02521-11	Sequence 11, Appli
10	51.5	18.5	341	4	US-09-008-465-1	Sequence 1, Appli
11	51.5	18.5	341	4	US-09-528-959-1	Sequence 1, Appli
12	51	18.3	416	1	US-08-117-083-61	Sequence 1, Appli
13	51	18.3	543	4	US-09-362-123A-4	Sequence 61, Appli
14	50.5	18.1	314	1	US-08-269-441A-2	Sequence 4, Appli
15	50.5	18.1	314	4	US-08-008-465-3	Sequence 2, Appli
16	50.5	18.1	314	4	US-09-276-851-2	Sequence 3, Appli
17	50.5	18.1	314	4	US-09-528-959-3	Sequence 2, Appli
18	50.5	18.1	381	3	US-08-911-853-27	Sequence 3, Appli
19	50.5	18.1	381	4	US-09-479-409-27	Sequence 27, Appli
20	50.5	18.1	381	4	US-09-479-453-27	Sequence 27, Appli
21	50	17.9	664	2	US-08-852-153-8	Sequence 27, Appli
22	50	17.9	1068	3	US-08-390-874C-11	Sequence 8, Appli
23	50	17.9	1068	4	US-09-265-772-11	Sequence 11, Appli
24	50	17.9	1069	2	US-08-162-081B-37	Sequence 11, Appli
25	50	17.9	1069	2	US-08-780-872-37	Sequence 37, Appli
26	50	17.9	1069	4	US-09-085-957-37	Sequence 37, Appli
27	50	17.9	1080	2	US-08-162-081B-36	Sequence 36, Appli

APPLICANT: Christensen, Tove  
APPLICANT: Lehbeck, Jan  
TITLE OF INVENTION: A Fungus Wherein The areaA, pepC and/or  
TITLE OF INVENTION: Pepe Genes Have Been Inactivated  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 60134520 No. 6013452disk of No. 6013452th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,415  
FILING DATE: 14-MAY-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 4657.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 866 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-079-415-6  
Query Match 92.5%; Score 258; DB 3; Length 866;  
Best Local Similarity 92.0%; Pred. No. 8.7e-29;  
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTNCGTKTTPLWRNPQGOPLCNACGLFLKHLGVVRPLSLKTDVIKKRQ 50  
DB 664 CTNCFQTPTPLWRNPQGOPLCNACGLFLKHLGVVRPLSLKTDVIKKR 713  
RESULT 3  
US-08-750-458A-2  
Sequence 2, Application US/08750458A  
Patent No. 6025185  
GENERAL INFORMATION:  
APPLICANT: Christensen, Tove  
APPLICANT: Hynes, Michael J.  
TITLE OF INVENTION: A Fungus Wherein The Area Gene Has Been Modified And An Area  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 60251850 No. 6025185disk of No. 6025185th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/750,458A  
FILING DATE: 3-December-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol E.  
REGISTRATION NUMBER: 36,993

APPLICANT: Christensen, Tove  
APPLICANT: Lehbeck, Jan  
TITLE OF INVENTION: A Fungus Wherein The areaA, pepC and/or  
TITLE OF INVENTION: Pepe Genes Have Been Inactivated  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 60134520 No. 6013452disk of No. 6013452th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10174  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,415  
FILING DATE: 14-MAY-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol  
REGISTRATION NUMBER: 36,993  
REFERENCE/DOCKET NUMBER: 4657.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 866 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-079-415-6  
Query Match 92.5%; Score 258; DB 3; Length 866;  
Best Local Similarity 92.0%; Pred. No. 8.7e-29;  
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTNCGTKTTPLWRNPQGOPLCNACGLFLKHLGVVRPLSLKTDVIKKRQ 50  
DB 664 CTNCFQTPTPLWRNPQGOPLCNACGLFLKHLGVVRPLSLKTDVIKKR 713  
RESULT 3  
US-08-750-458A-2  
Sequence 2, Application US/08750458A  
Patent No. 6025185  
GENERAL INFORMATION:  
APPLICANT: Christensen, Tove  
APPLICANT: Hynes, Michael J.  
TITLE OF INVENTION: A Fungus Wherein The Area Gene Has Been Modified And An Area  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 60251850 No. 6025185disk of No. 6025185th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6400  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/750,458A  
FILING DATE: 3-December-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Rozek, Carol E.  
REGISTRATION NUMBER: 36,993

REFERENCE/DOCKET NUMBER: 4129.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 866 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-750-458A-2  
Query Match 92.5%; Score 258; DB 3; Length 866;  
Best Local Similarity 92.0%; Pred. No. 8.7e-29;  
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTNCGTKTTPLWRNPQGOPLCNACGLFLKHLGVVRPLSLKTDVIKKR 50  
DB 664 CTNCFQTPTPLWRNPQGOPLCNACGLFLKHLGVVRPLSLKTDVIKKR 713  
RESULT 4  
US-08-927-394-2  
Sequence 2, Application US/08927394  
Patent No. 5990092  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: GATA-6 TRANSCRIPTION FACTOR: COMPOSITIONS AND METHODS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/927,394  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/025,574  
FILING DATE: 06-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: S1237/7005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 449 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-927-394-2  
Query Match 67.4%; Score 188; DB 2; Length 449;  
Best Local Similarity 62.0%; Pred. No. 5e-19;  
Matches 31; Conservative 10; Mismatches 9; Indels 0; Gaps 0;  
QY 1 CTNCGTKTTPLWRNPQGOPLCNACGLFLKHLGVVRPLSLKTDVIKKR 50  
DB 298 CANCHTTTTLWRNAGEPVCNACGLYMKLHGVPRPLAMKKEGIQTRK 347  
RESULT 5

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:42:57 ; Search time 48 Seconds  
(without alignments)  
214.632 Million cell updates/sec

Title: US-09-725-010-2  
 Perfect score: 279  
 Sequence: 1 CFNCGTKTTPLWRNPQGP.....LHGVRPLSEKTDVKKRQR 50

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

```
Searched:      671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters:  671580
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :
SPTRMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvcluss.*
16: sp_bacteriap.*
17: sp_archaeap.*
```

Pred! No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	%			SUMMARIES			Description	
	Score	Match	Length	DB	ID			
1	258	92.5	853	3	09C414		09c414	aspergillus
2	258	92.5	866	3	09V788		09v7e8	aspergillus
3	258	92.5	918	3	09HEW7		09hew7	cladosporiu
4	258	92.5	944	3	06C0043		06c0043	metarhizium
5	248	88.9	327	3	P870116		P870116	zygosacchar
6	242	86.7	1075	3	09P952		09p952	schizophyil
7	209	74.9	298	3	059842		059842	penicillium
8	203	72.8	312	3	09HEV2		09hev2	emericeila
9	203	72.8	320	3	09HEV3		09hev3	emericeila
10	203	72.8	436	3	09HEV4		09hev4	emericeila
11	201	72.0	324	5	09NJ18		09nj18	ceratitius c
12	201	72.0	511	5	08WQ04		08wq04	calliphora
13	199	71.3	537	5	095VY5		095vy5	anopheles g
14	198	71.0	486	5	08T3J0		08t3j0	drosofila
15	195	69.9	383	13	Q9W6U0		Q9w6u0	brachydanio
16	191	68.5	949	5	094884		094884	drosofila

RESULT 1  
09C414

ID	Q9C414	PRELIMINARY;	PRT;	853 AA.
DC	Q9C414			
AC	Q9C414.1			
DT	01-JUN-2001	(TrEMBLrel. 17, Created)		
DC	01-JUN-2001	(TrEMBLrel. 17, Last sequence update)		
DT	01-DEC-2001	(TrEMBLrel. 19, Last annotation update)		
DE	Area.			
DN	AREA.			
OS	Aspergillus oryzae.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurot			
CC	Eurotiales; Trichomaceae; mitosporic Trichomaceae			
OX	NCBI_TaxID=5062;			
XP	[1]			
RP	SEQUENCE FROM N.A.			
RA	STRAIN-TK3;			
RC	van den Broek P.;			
RT	"The A. oryzae areaA gene as a tool to overproduce a p			
RT	roteases";			
RL	Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databa			
DR	EMBL; AF320305; AAK08066.1; -;			
DR	HSP; P17429; 4GAT.			
DR	InterPro; IPR000679; Znf_GATA.			
DR	Pfam; PF003320; GATA; 1.			
DR	PRINTS; PR00619; GATAZNFINGER.			
DR	SMART; SM00401; Znf_GATA; 1.			
DR	PROSITE; PS00344; GATA_ZN_FINGER_1; 1.			
DR	PROSITE; PS50114; GATA_ZN_FINGER_2; 1.			
SQ	SEQUENCE 853 AA; 91497 MW; A5EC81CE256AB15 CRC			

## ALIGNMENTS

ID	Q9HEW7	PRELIMINARY;	PRT;	918 AA.
AC	Q9HEW7;			
DT	01-MAR-2001	(TREMBLrel. 16, Created)		
DT	01-WAR-2001	(TREMBLrel. 16, Last sequence update)		
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
DE	Nitrogen response factor NRF1			
DE	Cladosporium fulvum (Fulvia fulva).			
OS	Eukaryota; Fungi; Ascomycota; Pezizomycotina;			
OC	Dothideomycetes et Chaetothyriomycetes incertae sedis;			
OC	Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.			
OX	NCBI_TaxID=5499;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RP	Perez-Garcia A., Snoeijers S.S., Joosten M.H.A.J., Goosen T.,			
RA	De Wit P.J.G.M.;			
RA	"Expression of the avirulence gene Avr9 of the fungal tomato pathogen Cladosporium fulvum is regulated by the global nitrogen response factor NRF1."			
RT	Mol. Plant Microbe Interact. 0:0-0(2001).			
RL	EMBL; AF312694; AAG48616.1; -.			
DR	HSSP; PI7429; 4GAT.			
DR	InterPro; IPR002965; P-rich_extensn.			
DR	InterPro; IPR000679; znf_GATA.			
DR	Pfam; PF00320; GATA; 1.			
DR	PRINTS; PR00619; GATAZNFINGER.			
DR	PRINTS; PR01217; PRICHEXTENS.			
DR	SMART; SM00401; znf_GATA; 1.			
DR	PROSITE; PS00344; GATA_ZN_FINGER_1; 1.			
DR	PROSITE; PS50114; GATA_ZN_FINGER_2; 1.			
DR	SEQUENCE 918 AA; 99077 MW; E98198D999BCA899			
SQ	CRC64;			

8P7016	PRELIMINARY;	PRT;	327 AA.
AD	8P7016		
AC	8P7016;		
DT	01-JUL-1997 (TRENBLrel. 04, Created)		
DT	01-JUL-1997 (TRENBLrel. 04, Last sequence update)		
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)		
DE	DNA-binding protein.		
GN	SAT1.		
OS	Zygosaccharomyces rouxii (Candida mogii).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.		
NCBI_TaxId=4956;			
[1]			
SEQUENCE FROM N.A.			
USHU K., Otsuka H., Yoshikawa S., Taguchi G., Shimosaka M.,			
Mitsui N., Okazaki M.;			
"Cloning of the SAT1 gene concerned with salt tolerance of the yeast			
Zygosaccharomyces rouxii";			
J. ferment. Bioeng. 82:16-21(1996).			
EMBL; D83211; BAA18948.1; -.			
HSP; P17429; 4GAT.			
InterPro; IPR000679; Znf_GATA.			
Pfam; PF00320; GATA; 1.			





OC Eurotiales; Trichocomaceae; Emericella.  
OX NCBI\_TaxID=5072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-89218944; PubMed=2651886;  
RA Arst H.N. Jr., Tollervey D., Caddick M.X.;  
RD "A translocation associated, loss-of-function mutation in the nitrogen  
RE metabolite repression regulatory gene of Aspergillus nidulans can  
RF revert intracistronically."  
RG Mol. Gen. Genet. 215:364-367(1989).  
RH [2]  
RI SEQUENCE FROM N.A.  
RJ MEDLINE-21206224; PubMed=11309119;  
RK Conlon H., Zadra I., Haas H., Arst H.N. Jr., Jones M.G., Caddick M.X.;  
RL "The Aspergillus nidulans GATA transcription factor gene areB encodes  
RM at least three proteins and features three classes of mutation.";  
RN Mol. Microbiol. 40:361-375(2001).  
RO EMBL; AF320976; AAG49353.1; -.  
RP HSSP; P17429; 4GAT.  
RQ InterPro; IPRO00679; ZnF\_GATA.  
RS Pfam; PF00320; GATA; 1.  
RT PRINTS; PR00619; GATAZNFINGER.  
RU SMART; SM00401; ZnF\_GATA; 1.  
RV PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
RW PROSITE; PS01114; GATA\_ZN\_FINGER\_2; 1.  
RX SEQUENCE 320 AA; 34765 MW; C625ED939E74DF5 CRC64;

Query Match 72.8%; Score 203; DB 3; Length 320;  
Best Local Similarity 74.0%; Pred. No. 2.2e-19;  
Matches 37; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

OY 1 CTNCGGTKTTPLRNRNPQGOLCNACGLFLKHLGVVVRPLSLKTDVIKKQR 50  
I IIII IIIII : I IIIIIIIIIII IIIIIIIII I  
DB 25 CQCNGTSKTPLWRRDELGSVLNACGLFLKHLGRPRPISLKTDVKSRR 74

RESULT 10  
OYHEV4 PRELIMINARY; PRT; 436 AA.  
ID O9HEV4;  
AC Q9HEV4;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE GATA factor AREB gamma (Fragment).  
OS Emericella nidulans (Aspergillus nidulans).  
GN Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; Emericella.  
OX NCBI\_TaxID=5072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-89218944; PubMed=2651886;  
RA Arst H.N. Jr., Tollervey D., Caddick M.X.;  
RD "A translocation associated, loss-of-function mutation in the nitrogen  
RE metabolite repression regulatory gene of Aspergillus nidulans can  
RF revert intracistronically."  
RG Mol. Gen. Genet. 215:364-367(1989).  
RH [2]  
RI SEQUENCE FROM N.A.  
RJ MEDLINE-21206224; PubMed=11309119;  
RK Conlon H., Zadra I., Haas H., Arst H.N. Jr., Jones M.G., Caddick M.X.;  
RL "The Aspergillus nidulans GATA transcription factor gene areB encodes  
RM at least three proteins and features three classes of mutation.";  
RN Mol. Microbiol. 40:361-375(2001).  
RO EMBL; AF320976; AAG49353.1; -.  
RP HSSP; P17429; 4GAT.  
RQ InterPro; IPRO00679; ZnF\_GATA.  
RS Pfam; PF00320; GATA; 1.  
RT PRINTS; PR00619; GATAZNFINGER.  
RU SMART; SM00401; ZnF\_GATA; 1.  
RV PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
RW PROSITE; PS01114; GATA\_ZN\_FINGER\_2; 1.  
RX NON\_TER 1  
FT

SQ SEQUENCE 436 AA; 46251 MW; F91457AE2F8F7BB6 CRC64;  
Query Match 72.8%; Score 203; DB 3; Length 436;  
Best Local Similarity 74.0%; Pred. No. 3e-19;  
Matches 37; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

OY 1 CTNCGGTKTTPLRNRNPQGOLCNACGLFLKHLGVVVRPLSLKTDVIKKQR 50  
I IIII IIIII : I IIIIIIIIIII IIIIIIIII I  
DB 141 CQCNGTSKTPLWRRDELGSVLNACGLFLKHLGRPRPISLKTDVKSRR 190

RESULT 11  
OYJ18 PRELIMINARY; PRT; 324 AA.  
ID Q9NJ18;  
AC Q9NJ18;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Zinc finger DNA binding protein pannier.  
GN PNR.  
OS Ceratitis capitata (Mediterranean fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Tephritidae; Tephritidae; Ceratitis.  
OX NCBI\_TaxID=72113;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capitata, suggests a  
common origin for the stereotyped bristle patterns of higher  
Diptera";  
RL Development 127:1411-1420(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BENAKEO;  
RX MEDLINE-20171265; PubMed=10704387;  
RA Wulbeck C., Simpson P.;  
RT "Expression of achaete-scute homologues in discrete proneural clusters  
on the developing notum of the medfly Ceratitis capit

OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RP Stapleton M., Brokstein P., Hong L., Aghavani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas A., Park S.,  
RA Patel S., Phouanetavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Celniker S.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AY095095; AA011423.1; -.  
SQ SEQUENCE 486 AA; 51606 MW; BC0E10E9B8D4A0B CRC64;

Query Match 71.0%; Score 198; DB 5; Length 486;  
Best Local Similarity 66.0%; Pred. No. 1.6e-18;  
Matches 33; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

**Qy** 1 CTNCGTKTPTLWRRNPQGOPLCNACGLFCLKHGVRPLSLKTDVIKKRQR 50  
|||||: ||||| :||:|||||: ||||| ||:: | : | : |  
**Dd** 174 CTNCGTRTTTLWRRNRNDGPEVCNACGLYYKLHGWRPLAMRKDGIOIKR 223

RESULT	15
Q9W6U0	
ID	Q9W6U0 PRELIMINARY; PRT; 383 AA.
AC	Q9W6U0;
DT	01-NOV-1999 (TEMBLrel. 12, Created)
DT	01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT	01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE	gata5,
DE	GATA5 OR GTA5.
GN	Carchydanio rerio (Zebrafish) (Zebra danio).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC	Cyprinidae; Danio.
OX	NCBI_TaxID=7955;
[1]	
RN	SEQUENCE FROM N.A.
RP	

RP	SEQUENCE FROM N.A.
RC	TISSUE=EMBRYO;
RX	MEDLINE=99307072; PubMed=10375499;
RA	Rodaway A., Takeda H., Koshida S., Broadbent J., Price B.M.J.,
RA	Smith J.C., Patient R., Holder N.;
RT	Induction of the mesoderm in the zebrafish germ ring by yolk cell
RT	derived TGF-(beta) family signals and discrimination of mesoderm and
RT	endoderm by FGF.;
RL	Development 126:3067-3077(1999).
RL	EMBL: AJ242515; CAB43400.1; -.
DR	HSSP: P17679; 1GNF.
DR	2FIN; ZDB-GENE=980526-340; gata5.
DR	InterPro: IPR001164; hrip_like.
DR	InterPro: IPR000679; znf_GATA.
DR	Pfam: PF00320; GATA; 2.
DR	PRINTS: PR00619; GATAZNFINGER.
DR	SMART: SM00401; znf_GATA; 2.
DR	PROSITE: PS00344; GATA_ZN_FINGER_1; 2.
DR	PROSITE: PS0114; GATA_ZN_FINGER_2; 1.
SQ	SEQUENCE 383 AA; 41479 MW; 7F752664185588AF CRC64;

```

InterPro: IPR000679; Znf_GATA.
Pfam: PF00320; GATA: 2.
PRINTS: PR00619; GATAZNFINGER.
SMART: SM00401; Znf.GATA: 2.
PROSITE: PS00344; GATA_ZN_FINGER_1: 2.
PROSITE: PS50114; GATA_ZN_FINGER_2: 1.
SEQUENCE 383 AA; 41479 MW; 7F752664185588AF CRC64;

Query Match          69.9%  score 195;  DB 13;  Length 383;
Best Local Similarity 64.0%  Pred. No. 3.2e-18;
Matches 32;  Conservative 10;  Mismatches 8;  Indels

```

DR . PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 2.  
DR PROSITE; PS01114; GATA\_ZN\_FINGER\_2; 1.  
SQ SEQUENCE 383 AA; 41479 MW; 7F7526664185588AF CRC64;

```
Query Match          69.9%; Score 195; DB 13; Length 383;
Best Local Similarity 64.0%; Pred. No. 3,2e-18;
Matches 32; Conservative 10; Mismatches 8; Indels 0; Gaps 0;
```

QY 1 CTNCGTKTTPHRRNPQGQLCNACGLFLKLHGVRPLSLKTDVKKRQR 50  
||||| || ||||| :||:||||||:||||| :||: | :|  
Db 241 CTCNCHSTTTLWRNBAECPVCNAGLYMKLHGVPRLAMKESIQTRKR 290

Search completed: January 3, 2003, 19:50:34  
Job time : 49 secs

Search completed: January 3, 2003, 19:50:34  
Job time : 49 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 18:50:12 ; Search time 14 Seconds  
(without alignments)  
148.130 Million cell updates/sec

Title: US-09-725-010-3  
Perfect score: 281  
Sequence: 1 CTNCFQTPTPLRRNPDCQP.....LHGVRPLSLKTDVKKRNR 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	281	100.0	1036	1	NIT2_NEUCR
2	277	98.6	725	1	AREA_PENCH
3	277	98.6	860	1	AREA_PENRO
4	277	98.6	865	1	NREA_PENUR
5	277	98.6	866	1	AREA_ASPOR
6	277	98.6	876	1	AREA_EMENI
7	277	98.6	882	1	AREA_ASPNG
8	277	98.6	971	1	AREA_GIBFU
9	269	95.7	956	1	NUT1_MAGGR
10	252	89.7	855	1	GAF1_SCHPO
11	231	82.2	510	1	GAT1_YEAST
12	211	75.1	426	1	ASB4_NEUCR
13	210	74.7	551	1	GZF3_YEAST
14	204	72.6	269	1	DA80_YEAST
15	201	71.5	730	1	GLN3_YEAST
16	195	69.4	388	1	GASB_XENLA
17	195	69.4	390	1	GASA_XENLA
18	193	68.7	540	1	PNR_DROME
19	190	67.6	404	1	GAT5_MOUSE
20	190	67.6	779	1	SRP_DROME
21	189	67.3	391	1	GAGA_XENLA
22	189	67.3	391	1	GAT5_CHICK
23	189	67.3	397	1	GAT5_HUMAN
24	188	66.9	387	1	GAT6_CHICK
25	188	66.9	391	1	GAT6_XENLA
26	188	66.9	392	1	GAT4_XENLA
27	188	66.9	441	1	GAT6_RAT
28	188	66.9	444	1	GAT6_MOUSE
29	188	66.9	449	1	GAT6_HUMAN
30	187	66.5	950	1	URB1_USTMA
31	185	65.8	380	1	GAT4_CHICK
32	183	65.1	413	1	GAT1_MOUSE
33	183	65.1	440	1	GAT4_MOUSE

34	183	65.1	440	1	GAT4_RAT
35	183	65.1	442	1	GAT4_HUMAN
36	182	64.8	359	1	GAT4_XENLA
37	182	64.8	413	1	GAT1_HUMAN
38	182	64.8	416	1	GAT1_RAT
39	182	64.8	416	1	ELT1_CAEEL
40	181	64.4	364	1	GAT1B_XENLA
41	181	64.4	564	1	GAF2_SCHPO
42	180	64.1	452	1	GAT2_XENLA
43	180	64.1	486	1	GATC_DROME
44	179	63.7	304	1	GAT1_CHICK
45	179	63.7	466	1	GAT2_CHICK

ALIGNMENTS

RESULT 1

NIT2\_NEUCR  
ID NIT2\_NEUCR STANDARD; PRT; 1036 AA.  
AC P19212;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nitrogen catabolic enzyme regulatory protein.  
GN NIT-2  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A;  
RX MEDLINE=90158568; PubMed=2137552;  
RA Fu Y.-H., Marzluf G.A.;  
RT "nit-2, the major nitrogen regulatory gene of Neurospora crassa,"  
RT encodes a protein with a putative zinc finger DNA-binding domain.";  
RL Mol. Cell. Biol. 10:1056-1065(1990).  
RN [2]  
RP MUTAGENESIS.  
RX MEDLINE=91186820; PubMed=2150539;  
RA Fu Y.-H., Marzluf G.A.;  
RT "Site-directed mutagenesis of the 'zinc finger' DNA-binding domain of the nitrogen-regulatory protein NIT2 of Neurospora.";  
RL Mol. Microbiol. 4:1847-1852(1990).  
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. DURING CONDITIONS OF NITROGEN LIMITATION IT TURNS ON THE EXPRESSION OF GENES FOR ENZYMES WHICH ARE REQUIRED FOR THE USE OF A VARIETY OF SECONDARY NITROGEN SOURCES, INCLUDING NITRATES, PURINES, AMINO ACIDS, AND PROTEINS.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- INDUCTION: BY LACK OF A PRIMARY NITROGEN SOURCE.  
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

CC EMBL; M33956; AAB03891.1; -  
CC PIR; A34755; A34755.  
CC HSSP; P17429; AGAT.  
CC TRANSPAC; T00627; -  
CC InterPro; IPR000679; Znf\_GATA.  
CC Pfam; PF00320; GATA; 1.  
CC PRINTS; PR00619; GATAZNFINGER.  
CC SMART; SM00401; Znf.GATA; 1.  
CC PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
CC PROSITE; PS50114; GATA\_ZN\_FINGER\_2; 1.  
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;

KW Nuclear protein; Nitrate assimilation; Repeat.  
FT DOMAIN 49 110 3 X APPROXIMATE REPEATS.  
FT REPEAT 49 55 1.  
FT REPEAT 87 92 2.  
FT REPEAT 105 110 3.  
FT ZN\_FING 743 767 GATA-TYPE.  
FT MUTAGEN 743 746 CTRC->STNG: ABOLISHES DNA-BINDING.  
FT MUTAGEN 755 756 RR->GG: ABOLISHES DNA-BINDING.  
FT MUTAGEN 765 766 NA->DV: ABOLISHES DNA-BINDING.  
FT MUTAGEN 768 769 GL->DV: ABOLISHES DNA-BINDING.  
FT MUTAGEN 789 790 KR->NS: ABOLISHES DNA-BINDING.  
SQ SEQUENCE 1036 AA; 109310 MW; 5FE4992B1C223514 CRC64;

Query Match 100.0%; Score 281; DB 1; Length 1036;  
Best Local Similarity 100.0%; Pred. No. 2.4e-28;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFCTOTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVVKRNR 50  
|||||  
DB 743 CTNCFCTOTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVVKRNR 792

RESULT 2  
AREA\_PENCH STANDARD; PRT; 725 AA.  
AC Q01582;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Nitrogen regulatory protein area (Nitrogen regulator nre).  
GN AREA OR NRE.  
OS Penicillium chrysogenum.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
OX NCBI\_TaxID=5076;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=95308537; PubMed=7788718;  
RA Haas H., Bauer B., Redl B., Stoeffler G., Marzluf G.A.;  
RT "Molecular cloning and analysis of nre, the major nitrogen regulatory  
gene of Penicillium chrysogenum.";  
RL Curr. Genet. 27:150-158(1995).  
CC 1- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN.  
CC 1- SUBCELLULAR LOCATION: Nuclear.  
CC 1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; U02612; AAA83400.1; -.  
DR HSP; P17429; 4GAT.  
DR InterPro: IPR000679; Znf\_GATA.  
DR Pfam: PF00320; GATA; 1.  
DR PRINTS; PRO0619; GATAZNFINGER.  
DR SMART; SM00401; Znf\_GATA; 1.  
DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
DR PROSITE; PS01114; GATA\_ZN\_FINGER\_2; 1.  
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;  
KW Nuclear protein; Nitrate assimilation.  
FT ZN\_FING 525 549 GATA-TYPE.  
SQ SEQUENCE 725 AA; 76848 MW; 45C24148F81F8D43 CRC64;

Query Match 98.6%; Score 277; DB 1; Length 725;  
Best Local Similarity 98.0%; Pred. No. 5.7e-28;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFCTOTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVVKRNR 50  
|||||

DB 525 CTNCFCTOTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVVKRNR 574

## RESULT 3

AREA\_PENRO STANDARD; PRT; 860 AA.  
AC Q13508;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nitrogen regulatory protein area (Nitrogen regulator nmc).  
GN AREA OR NMC.  
OS Penicillium roqueforti.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
OX NCBI\_TaxID=5082;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=P2;

RA Gente S., Poussereau N., Fevre M.;

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

CC 1- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN.

CC 1- SUBCELLULAR LOCATION: Nuclear.

CC 1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AJ001530; CAA04815.1; -.

DR HSP; P17429; 4GAT

DR InterPro: IPR000679; Znf\_GATA.

DR Pfam: PF00320; GATA; 1.

DR PRINTS; PRO0619; GATAZNFINGER.

DR SMART; SM00401; Znf\_GATA; 1.

DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.

DR PROSITE; PS01114; GATA\_ZN\_FINGER\_2; 1.

KW Transcription regulation; Activator; DNA-binding; Zinc-finger;

KW Nuclear protein; Nitrate assimilation.

FT ZN\_FING 660 684 GATA-TYPE.

SQ SEQUENCE 860 AA; 91615 MW; 2B8BF0DDC1BA68 CRC64;

Query Match 98.6%; Score 277; DB 1; Length 860;  
Best Local Similarity 98.0%; Pred. No. 6.7e-28;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFCTOTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVVKRNR 50  
|||||

DB 660 CTNCFCTOTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVVKRNR 709

## RESULT 4

NRFA\_PENUR

ID NRFA\_PENUR STANDARD; PRT; 865 AA.

AC Q92269;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Nitrogen regulatory protein NRFA.

GN NRFA.

OS Penicillium urticae.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.

OX NCBI\_TaxID=29844;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL 2159A;

RA Ellis C.M.;

RL Thesis (1996), University of Calgary, Canada.

```
DR PROSITE; PS00114; GATA_ZN_FINGER_2; 1.  
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;  
KW Nuclear protein; Nitrate assimilation.  
FT ZN_FING      664    688     GATA-TYPE.  
SQ SEQUENCE   866 AA; 92882 MW; C09A03EB12E3FBB4 CRC64;
```

Query Match 98.6%; Score 277; DB 1; Length 866;  
Best Local Similarity 98.0%; Pred. No. 6,7e-28;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps

Qy 1 CTNCFTQTTPLLWRNPQGQLPCNACGLFLKLHG VVVRPLSLKTDVIKKRNR 50  
-----  
IDB 664 CTNCFTQTTPLLWRNPGEQPLCNAGLFLKLHG VVVRPLSLKTDVIKKRNR 1713

RESULT 6  
AREA\_EMENI STANDARD; PRT; 876 AA.

ID AREA EMENI AC PI7429;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Nitrogen regulatory protein area.  
GN AREA.  
OS Emericella nidulans (*Aspergillus nidulans*).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OX Eurotiales; Trichocomaceae; Emericella.  
OC NCBI\_TaxID=5072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90228331; PubMed=1970293;  
RA Kudla B., Caddick M.X., Langdon T., Martinez-Rossi N.M.,  
RA Bennett C.F., Sibley S., Davies R.W., Arst H.N. Jr.;  
PT "The regulatory gene areA mediating nitrogen metabolite repression in  
PT *Aspergillus nidulans*. Mutations affecting specificity of gene  
RT activation alter a loop residue of a putative zinc finger."  
RL EMBO J. 9:1355-1364(1990).  
RN [2]  
RP REVISIONS.  
RX MEDLINE=96123430; PubMed=8596437;  
RA Langdon T., Seerins A., Ravagnani A., Caddick M.X., Arst H.N. Jr.;  
PT "Mutational analysis reveals dispensability of the N-terminal region  
PT of the *Aspergillus* transcription factor mediating nitrogen metabolite  
RT repression.";  
RL Mol. Microbiol. 17:877-888(1995).  
RN [3]

RX STRUCTURE BY NMR OF 662-727.  
RX MEDLINE=98202574; PubMed=9533883;  
RA Starich M.R., Wikstroem M., Arst H.N. Jr., Clore G.M., Gronenborn A.M.;  
RA "The solution structure of a fungal ARE protein-DNA complex: an  
RT alternative binding mode for the basic carboxyl tail of GATA  
RT factors,"  
J. Mol. Biol. 277:605-620(1998).  
CC -! FUNCTION: TRANSCRIPTION ACTIVATOR THAT MEDIATES NITROGEN  
CC METABOLITE REPRESSON IN A NIDULANS. ACTIVATE THE TRANSCRIPTION  
CC OF UAPA.  
CC -! SUBCELLULAR LOCATION: Nuclear.  
CC -! SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce or send an email to license@isb-sib.ch).

EMBL; X52491; CAA36731.1; --  
PIR; S10017; SI0017.  
PDB; 4GAT; 28-JAN-98.  
PDB; 5GAT; 28-JAN-98.  
PDB; 6GAT; 28-JAN-98.

```
DR PDB: 7GAT; 28-JAN-98.
DR TRANSFAC; T02533; -.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation; 3D-structure.
FT ZN_FING 673 697 GATA-TYPE.
FT DNA_BIND 721 742 H-T-H MOTIF (PROBABLE).
SQ SEQUENCE 876 AA; 94195 MW; 9ADC2273EE536F98 CRC64;

Query Match 98.6%; Score 277; DB 1; Length 876;
Best Local Similarity 98.0%; Pred. No. 6.8e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGWVRPLSLKTDVIKKRNR 50
DB 673 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGWVRPLSLKTDVIKKRNR 722

RESULT 7
AREA ASPNG
ID AREA ASPNG STANDARD; PRT; 882 AA.
AC O13412;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N402;
RA Macabe A.P., Vanhanen S.A.S., Sollewijn Gelpke M.,
RA van de Vondervoort P., Arst H.N., Visser J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X81998; CAA57524.1; -.
DR HSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 676 700 GATA-TYPE.
SQ SEQUENCE 882 AA; 94518 MW; 97A502936B94E5AE CRC64;

Query Match 98.6%; Score 277; DB 1; Length 882;
Best Local Similarity 98.0%; Pred. No. 6.8e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGWVRPLSLKTDVIKKRNR 50
```

```
DB 676 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGWVRPLSLKTDVIKKRNR 725

RESULT 8
AREA GIBFU
ID AREA GIBFU STANDARD; PRT; 971 AA.
AC P78688;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Gibberella fujikuroi (Bakanae and foot rot disease fungus) (Fusarium
OS moniliforme).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=m567;
RA MEDLINE=99168774; PubMed=10071216;
RA tudzynski B., Homann V., Feng B., Marzluf G.A.;
RT "Isolation, characterization and disruption of the area nitrogen
RT regulatory gene of Gibberella fujikuroi.";
RL Mol. Gen. Genet. 261:106-114(1999).
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11006; CAA71897.1; -.
DR HSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 694 718 GATA-TYPE.
SQ SEQUENCE 971 AA; 103580 MW; 887DD882141C7453 CRC64;

Query Match 98.6%; Score 277; DB 1; Length 971;
Best Local Similarity 98.0%; Pred. No. 7.5e-28;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGWVRPLSLKTDVIKKRNR 50
DB 694 CTNCFQTTPPLWRNPDPGQPLCNACGLFLKLGWVRPLSLKTDVIKKRNR 743

RESULT 9
AREA MAGGR
ID AREA MAGGR STANDARD; PRT; 956 AA.
AC Q01168;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein NUT1.
GN NUT1.
OS Magnaporthe grisea (Rice blast fungus) (Pyricularia grisea).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
OC
```



```

OX NCBI_TaxID=148305;
RN [1]
RC SEQUENCE FROM N.A.
PC STRAIN=Guyane 11;
RX MEDLINE=96335139; PubMed=8757395;
RA Froeliger E.H., Carpenter B.E.;
RT "NUT1, a major nitrogen regulatory gene in Magnaporthe grisea, is
RT dispensable for pathogenicity.";
RL Mol. Gen. Genet. 251:647-656(1996).
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN; ACTIVATES EXPRESSION
CC OF NITROGEN-REGULATED GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U60290; AAB03415.1; -
CC HSSP: P17429; 4GAT.
CC TRANSFAC; T02829; -
CC InterPro; IPR000679; Znf_GATA.
CC PRINTS; PR00619; GATAZNFINGER.
CC SMART; SM00401; Znf_GATA; 1.
CC PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
CC PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation.
FT ZN_FING 663 687 GATA-TYPE.
SQ SEQUENCE 956 AA; 100874 MW; 40ABDASA07A7D7AB CRC64;

Query Match 95.7%; Score 269; DB 1; Length 956;
Best Local Similarity 96.0%; Pred. No. 7,9e-27;
Matches 48; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNCFQTPTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRR 50
DB 663 CTNCFQTPTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRR 712

RESULT 10
GAF1_SCHPO
ID GAF1_SCHPO STANDARD; PRT; 855 AA.
AC Q10280; O94482; O9USK9;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription factor gaf1 (Gaf-1).
GN GAF1 OR SPCC417.01C OR SPCC1902.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RC SEQUENCE FROM N.A.
PC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feitwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

```



```

zipper.":
RL
[4]
SEQUENCE FROM N.A.
Urrestarazu L.A., Jauniaux J.-C.;
Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: NEGATIVE REGULATOR OF MULTIPLE NITROGEN CATABOLIC
GENES INCLUDING THE ALLANTOIN PATHWAY GENES.
-!- SUBCELLULAR LOCATION: NUCLEAR
-!- INDUCTION: SENSITIVE TO NITROGEN CATABOLITE REPRESSION.
-!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
EMBL; M77821; AAA34556.1; -
EMBL; X60199; CAA42757.1; -
EMBL; Z28259; CAA82107.1; -
EMBL; Z28258; CAA82106.1; -
PIR; S22781; S22781.
HSSP; P17429; 4GAT.
TRANSFAC; T02411; -
SGD; S0001742; DAL80.
InterPro; IPR000679; Znf_GATA.
Pfam; PF003320; GATA; 1.
PRINTS; PR00619; GATAZNFINGER.
SMART; SM00401; Znf_GATA; 1.
PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
Transcription regulation; Repressor; DNA-binding; Zinc-finger;
Nuclear protein; Nitrate assimilation.
ZNF_FING 31 55 GATA-TYPE.
DOMAIN 79 100 ASN-RICH.
DOMAIN 101 108 ARG/LYS-RICH (BASIC).
CONFLICT 6 6 S -> L (IN REF. 3).
CONFLICT 207 207 V -> I (IN REF. 3).
SEQUENCE 269 AA; 30166 MW; F37BEE10038599EC CRC64;

Query Match 72.68; Score 204; DB 1; Length 269;
Best Local Similarity 74.08; Pred. No. 5.4e-19;
Matches 37; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 CINCFTQTTPLWRNPQGPLNCACGLFLKLHGVRPLSLKTDVIKKNR 50
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 31 CONCTVFTPLWRDEHGTVCNACGLFLKLHGEPRLSLKTDITKSNR 80
| | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
GLN3_YEAST
ID GLN3_YEAST STANDARD; PRT; 730 AA.
AC P18494;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein GLN3.
DE GLN3 OR YER040W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Ascomycotaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=92049353; PubMed=1682800;
RA Minehart P.L.; Nagasani B.;
RT "Sequence and expression of GLN3, a positive nitrogen regulatory gene
of Saccharomyces cerevisiae encoding a protein with a putative zinc
finger DNA-binding domain."
RL Mol. Cell. Biol. 11:6216-6228(1991).

```

[2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972;  
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,  
 RA Aviles E., Berne A., Brennan T., Carpenter J., Chen E., Cherry J.M.,  
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,  
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,  
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,  
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,  
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: POSITIVE NITROGEN REGULATORY PROTEIN. REQUIRED FOR  
 CC THE ACTIVATION OF TRANSCRIPTION OF A NUMBER OF GENES (INCLUDING  
 CC THE ALLANTOIN PATHWAY GENES) IN RESPONSE TO THE REPLACEMENT OF  
 CC GLUTAMINE BY GLUTAMATE AS SOURCE OF NITROGEN. BINDS THE NITROGEN  
 CC UPSTREAM ACTIVATION SEQUENCE OF GLN1, THE GENE ENCODING GLUTAMINE  
 CC SYNTHETASE. URE2 MAY CATALYTICALLY INACTIVATE GLN3 IN RESPONSE TO  
 CC AN INCREASE IN THE INTRACELLULAR CONCENTRATION OF GLUTAMINE..  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M35267; AAA34645.1; -;  
 DR EMBL; U18796; AAB64575.1; -;  
 DR PIR; S22280; S22280.  
 DR HSSP; PI7429; 4GAT.  
 DR TRANSFAC; T02818; -;  
 DR SGD; S0000842; GLN3.  
 DR InterPro; IPR000679; Znf\_GATA.  
 DR Pfam; PF00320; GATA.1.  
 DR PRINTS; PR00619; GATAZNFINGER.  
 DR SMART; SM00401; Znf\_GATA.1.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
 DR PROSITE; PS50114; GATA\_ZN\_FINGER\_2; 1.  
 KW Transcription regulation; Activator; DNA-binding; Zinc-finger;  
 KW Nuclear protein; Nitrate assimilation.  
 FT ZN\_FING 306 330 GATA-TYPE.  
 FT DOMAIN 351 361 ARG/LYS-RICH (BASIC).  
 FT CONFLICT 474 474 P -> G (IN REF. 1).  
 SQ SEQUENCE 730 AA; 79382 MW; 3159E1844469942E CRC64;  
 Query Match 71.5%; Score 201; DB 1; Length 730;  
 Best Local Similarity 75.0%; Pred. No. 3.3e-16;  
 Matches 36; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 CYNCFQTPTPLWRRNPDPGLNACGLFLKHLGVVRPLSLKTDVIKKR 48  
 Db 306 CFNCKTFTPLWRRSPGNTLCNACGLFQKLGHTNRPLSLKSDVIKKR 353

Search completed: January 3, 2003, 19:48:37  
 Job time : 14 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:46:47 ; Search time 25 seconds  
(without alignments)  
192.269 Million cell updates/sec

Title: US-09-725-010-3

Perfect score: 281

Sequence: 1 CINCFTQTTPLRNPDGQP.....LHGVRPLSLKTDVIKKNR 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	281	100.0	1036	1 A34755	nitrogen regulator
2	277	98.6	862	2 S51493	major nitrogen reg
3	277	98.6	876	1 A57988	regulatory protein
4	252	89.7	855	2 T41336	probable nitrogen
5	231	82.2	510	2 S56233	probable membrane
6	210	74.7	551	2 S53377	GZF3 protein - yea
7	204	72.6	269	2 S22781	transcription fact
8	201	71.5	730	2 S50543	GLN3 protein - yea
9	195	69.4	388	2 S51420	transcription fact
10	195	69.4	390	2 I51419	transcription fact
11	190	67.6	779	1 S40382	box A-binding fact
12	189	67.3	391	2 I50702	transcription fact
13	188	66.9	387	2 I50703	transcription fact
14	188	66.9	391	2 A49645	transcription fact
15	187	66.5	950	2 S27473	URB1 protein - sm
16	185	65.8	380	2 I50701	transcription fact
17	183	65.1	413	2 S04655	transcription fact
18	183	65.1	439	1 A48099	transcription fact
19	183	65.1	439	2 I57561	transcription fact
20	183	65.1	440	2 I61183	transcription fact
21	182	64.8	359	2 A41602	transcription fact
22	182	64.8	413	1 A34888	transcription fact
23	182	64.8	413	2 S48756	transcription fact
24	182	64.8	416	1 A41267	transcription fact
25	182	64.8	454	2 T26296	hypothetical prote
26	181	64.4	364	2 B41602	transcription fact
27	181	64.4	564	2 T38291	GATA-type transcri
28	180	64.1	452	2 C41602	transcription fact
29	180	64.1	486	1 A57601	transcription fact

30 179 63.7 118 2 B48099 transcription fact  
31 179 63.7 304 2 A32993 transcription fact  
32 179 63.7 466 2 A36389 transcription fact  
33 179 63.7 474 2 A41782 transcription fact  
34 179 63.7 480 2 A40815 GATA-transcription  
35 179 63.7 532 2 JG6170 transcription fact  
36 178 63.3 435 2 D41602 transcription fact  
37 178 63.3 443 2 B39794 transcription fact  
38 178 63.3 444 1 A39794 transcription fact  
39 178 63.3 444 2 B36389 transcription fact  
40 175 62.3 564 2 T43298 transcription fact  
41 171 60.9 241 2 S53812 BmGATA beta isofo  
42 171 60.9 327 2 S53811 BmGATA beta isofo  
43 171 60.9 509 2 A53741 transcription fact  
44 150.5 53.6 433 2 A56953 transcription fact  
45 150.5 53.6 613 2 T19677 hypothetical prote

## ALIGNMENTS

### RESULT 1

A34755

nitrogen regulatory protein nit-2 - Neurospora crassa

C:Species: Neurospora crassa

C>Date: 13-Jul-1990 #sequence\_revision 26-Jul-1996 #text\_change 16-Feb-1997

C:Accession: A34755

R:Fu, Y.H.; Marzluf, G.A.

Mol. Cell. Biol. 10, 1056-1065, 1990

A>Title: nit-2, the major nitrogen regulatory gene of Neurospora crassa, encodes a pr

A:Reference number: A34755; MUID:90158568; PMID:2137552

A:Accession: A34755

A:Molecule type: DNA; mRNA

A:Residues: 1-1036 <FOY>

A:Cross-references: GB:M33956

C:Genetics:

A:Introns: 209/2; 335/3

C:Superfamily: nitrogen regulatory protein nit-2; GATA-type zinc finger homology

C:Keywords: DNA binding; transcription regulation; zinc finger

F:740-793/Domain: GATA-type zinc finger homology <GZF>

F:743-767/Region: zinc finger GATA motif

Query Match

Best Local Similarity 100.0%; Score 281; DB 1; Length 1036;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CINCFTQTTPLRNPDGQPLCNACGLFLKLGVRPLSLKTDVIKKNR 50

|||||

Db 743 CINCFTQTTPLRNPDGQPLCNACGLFLKLGVRPLSLKTDVIKKNR 792

### RESULT 2

S51493

major nitrogen regulation protein - Penicillium chrysogenum (strain Q176)

C:Species: Penicillium chrysogenum

A:Variety: strain Q176

C>Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000

C:Accession: S51493

R:Haas, H.; Bauer, B.; Stoeffler, G.; Marzluf, G.A.

Curr. Genet. 27, 150-158, 1995

A>Title: Molecular cloning and analysis of nre, the major nitrogen regulatory gene of

A:Reference number: S51493; MUID:95308537; PMID:7798718

A:Accession: S51493

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-862 <HAA>

A:Cross-references: EMBL:U02612

C:Genetics:

A:Introns: 123/2

C:Superfamily: nitrogen regulatory protein nit-2; GATA-type zinc finger homology

C:Keywords: zinc finger

F:659-712/Domain: GATA-type zinc finger homology <GZF>



A:Molecule type: DNA  
A:Residues: 1-269 <CUN>  
A:Cross-references: EMBL:W7821; NID:g171371; PIDN:AAA34556.1; PID:g171372  
R:Urretazu, L.A.; Jauniaux, J.C.  
Submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38097  
A:Accession: S38106  
A:Molecule type: DNA  
A:Residues: 1-269 <URD>  
A:Cross-references: EMBL:Z28259; NID:g486468; PID:g486469; MIPS:YKR034w  
A:Experimental source: strain S288C  
R:Coornaert, D.; Vissers, S.; Andre, B.; Grensou, M.  
Curr. Genet. 21, 301-307, 1992  
A:title: The UGA43 negative regulatory gene of Saccharomyces cerevisiae contains both  
A:Reference number: S26868; MUID:92405246; PMID:1525858  
A:Accession: S26868  
A:Molecule type: DNA  
A:Residues: 1-5,'L','7'-206,'I','208-269 <COO>  
A:Cross-references: EMBL:X60199; NID:g4751; PIDN:CAA42757.1; PID:g4752  
C:Genetics:  
A:Gene: SGD:DAL80; UGA43  
A:Cross-references: SGD:S0001742; MIPS:YKR034w  
A:Map position: 11R  
C:Superfamily: GATA-type zinc finger homology  
C:Keywords: DNA binding; nucleus; transcription factor; zinc finger  
F:28-81/Domain: GATA-type zinc finger homology <GZF>  
F:31-55/Region: zinc finger GATA motif

Query Match 72.6%; Score 204; DB 2; Length 269;  
Best Local Similarity 74.0%; Pred. No. 2e-18;  
Matches 37; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 CTNCFTTPTPLWRNPDPGLCNACGLFLKHLGVVRPLSLKTVDVIKKNR 50  
| | | | | | | | | | | | | | | | | | | | | |  
Db 31 CQCFTVTKPLWRDEHGTCLNACGLFLKHLGEPRPISLKTDTIKSRNR 80

RESULT 8  
S50543  
GLN3 protein - yeast (Saccharomycetes cerevisiae)  
N:Alternate names: protein YER040w  
C:Species: Saccharomycetes cerevisiae  
C>Date: 28-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999  
C:Accession: S50543; S22280  
R:Dietrich, F.S.  
submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of S. cerevisiae cosmid 9379, 9581, and lambda clone 467  
A:Reference number: S50536  
A:Accession: S50543  
A:Molecule type: DNA  
A:Residues: 1-730 <DIE>  
A:Cross-references: EMBL:U18796; NID:g603265; PIDN:AAB64575.1; PID:g603273; MIPS:YER040w  
R:Minehart, P.L.; Magasanik, B.  
Mol. Cell. Biol. 11, 6216-6228, 1991  
A:title: Sequence and expression of GLN3, a positive nitrogen regulatory gene of Sacc  
A:Reference number: S22280; MUID:92049353; PMID:1692800  
A:Accession: S22280  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-473,'G','475-730 <MIN>  
C:Genetics:  
A:Gene: SGD:GLN3  
A:Cross-references: SGD:S0000842; MIPS:YER040w  
A:Map position: 5R  
C:Superfamily: GATA-type zinc finger homology  
C:Keywords: zinc finger  
F:303-356/Domain: GATA-type zinc finger homology <GZF>  
F:306-330/Region: zinc finger GATA motif

Query Match 71.5%; Score 201; DB 2; Length 730;  
Best Local Similarity 75.0%; Pred. No. 1.2e-17;  
Matches 36; Conservative 4; Mismatches 8; Indels 0; Gaps 0;





```
QY      1 CTNCQTTPPLWRBNPDGQLCNACGLFLKLGHVVRPLSLKTDTVTKKNR 50
|||||
|||||
Db      482 CTNCQTTPPLWRDEDNCCNACGLHYHLGHTRPCLGMKSKSVTKRRKR 531
```

Search completed: January 3, 2003, 19:51:45  
Job time : 25 secs

Search completed: January 3, 2003, 19:51:45  
Job time : 25 secs

QY	1	235
Dd	CTNCFTQTTPLRRNPDDGQPLCNACGLFLKLHGVRPLSLKTDVYKKRNR	CANCHTTTTLRRNAEGEPVCNACGLYNKLHGVRPLAMKKEGIQTRKR

RESULT 14  
 A#9645 transcription factor GATA - rat  
 N:Alternate names: DNA binding protein  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 08-Dec-2000  
 C:Accession: A#9645  
 R:Tamura, S.; Wang, X.H.; Maeda, M.; Futai, M.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 10876-10880, 1993  
 A:Title: Gastric DNA-binding proteins recognize upstream sequence motifs of parietal cell  
 A:Reference number: A#9645; MUID:194068504; PMID:8248184

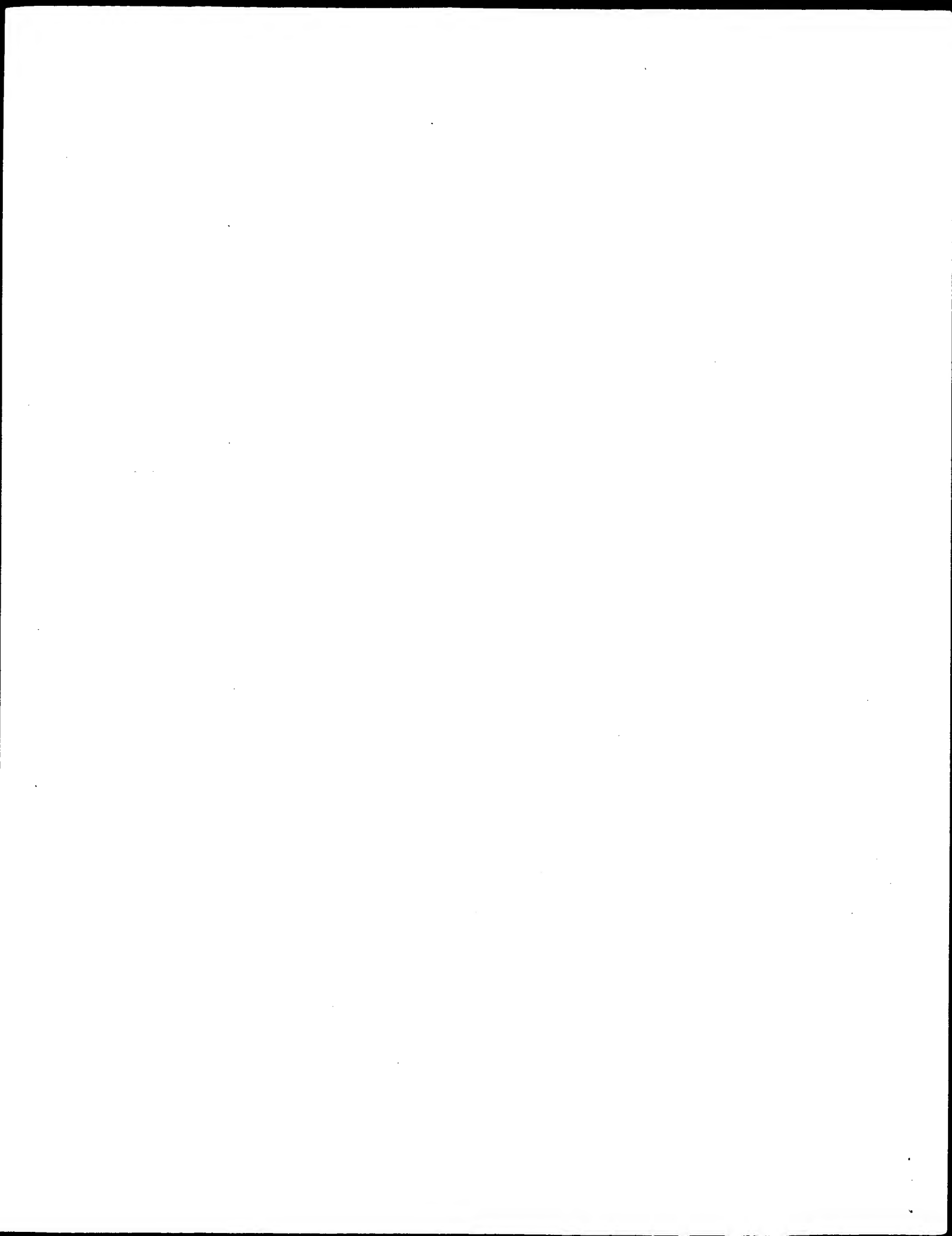
A:Accession: AY9643  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-391 <TAM>  
A:Cross-references: GB:L22760; NID:g437667; PID:g437668  
C:Genetics:  
A:Gene: GATA-GT1  
C:Superfamily: transcription factor GATA-4; GATA-type zinc finger  
C:Keywords: zinc finger  
F:185-238/Domain: GATA-type zinc finger homology <GZF1>  
F:239-292/Domain: GATA-type zinc finger homology <GZF2>

Query Match 66.9%; Score 188; DB 2; Length 391;  
Best Local Similarity 62.0%; Pred. No. 3.1e-16;  
Matches 31; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

Qy	1	CNCFQTTPLRNRNDQQPLCNACGLFLKLHGVRPLSLKTDVVKRNR	50
		:	
Db	242	CANCHTTTTLWRRNAEGPVCNACGLYMKLHGVPRPLAMKKEGIOTKR	291

RESULT 15  
S27473  
URBS1 protein - smut fungus (*Ustilago maydis*)  
C:Species: *Ustilago maydis* (corn smut)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Sep-1997  
C:Accession: S27473  
R:Voisard, C. P.; Wang, J.; Xu, P.; Leong, S. A.  
submitted to the EMBL Data Library, January 1992  
A:Description: Isolation and characterization of urbs1, a gene involved in the regulation of  
A:Reference number: S27473  
A:Accession: S27473  
A:Molecule type: DNA  
A:Residues: 1-950 <VOI>  
A:Cross-references: EMBL:W80547; NID:G295416; PID:g170596  
C:Superfamily: GATA-type zinc finger homology  
C:Keywords: DNA binding; transcription regulation; zinc finger  
F:335-388/Domain: GATA-type zinc finger homology <GZF1>  
F:479-532/Domain: GATA-type zinc finger homology <GZF2>

Query Match 66.5%; Score 187; DB 2; Length 950;  
Best Local Similarity 64.0%; Pred. No. 9.5e-16;  
Matches 32; Conservative 6; Mismatches 12; Indels



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:48:43 ; Search time 187.5 seconds  
(without alignments)  
171.929 Million cell updates/sec

Title: US-09-725-010-3  
Perfect score: 281  
Sequence: 1 CTNCTQTTPPLWRRNPDPGQ.....LHGVRPLSLKTDVTKRNR 50

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US09\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	281	100.0	1036	21	US-09-791-537-26785
2	281	100.0	1036	21	US-09-791-537-121036
3	277	98.6	66	21	US-09-791-537-44327
4	277	98.6	725	21	US-09-791-537-118817
5	277	98.6	860	21	US-09-791-537-125732
6	277	98.6	862	21	US-09-791-537-116651

Sequence 26785, A  
Sequence 121036, A  
Sequence 44327, A  
Sequence 118817, A  
Sequence 125732, A  
Sequence 116651, A

7	277	98.6	865	21	US-09-791-537-118818	Sequence 118818, A
8	277	98.6	866	11	US-08-750-458-2	Sequence 2, Appl1
9	277	98.6	866	21	US-09-791-537-44920	Sequence 44920, A
10	277	98.6	866	21	US-09-791-537-81933	Sequence 81933, A
11	277	98.6	876	18	US-09-487-558-44	Sequence 44, Appl
12	277	98.6	876	18	US-09-487-558B-44	Sequence 44, Appl
13	277	98.6	876	21	US-09-791-537-17502	Sequence 17502, A
14	277	98.6	876	21	US-09-791-537-47207	Sequence 47207, A
15	277	98.6	876	22	US-09-801-368-44	Sequence 44, Appl
16	277	98.6	882	21	US-09-791-537-81931	Sequence 81931, A
17	277	98.6	944	21	US-09-791-537-137385	Sequence 137385, A
18	277	98.6	971	21	US-09-791-537-151258	Sequence 151258, A
19	274	97.5	66	21	US-09-791-537-43833	Sequence 43833, A
20	269	95.7	956	21	US-09-791-537-118828	Sequence 118828, A
21	256	91.1	688	24	US-10-032-585-7876	Sequence 7876, Ap
22	256	91.1	719	15	US-09-248-796-17559	Sequence 17559, A
23	256	91.1	719	25	US-10-179-131-7398	Sequence 7398, Ap
24	256	91.1	719	27	US-60-096-409-17559	Sequence 17559, A
25	252	89.7	290	21	US-09-791-537-118820	Sequence 118820, A
26	252	89.7	727	21	US-09-791-537-148143	Sequence 148143, A
27	231	82.2	156	21	US-09-791-537-89028	Sequence 89028, A
28	231	82.2	510	21	US-09-791-537-118210	Sequence 118210, A
29	212	75.4	205	16	US-09-248-796-17564	Sequence 17564, A
30	212	75.4	205	27	US-60-096-409-17564	Sequence 17564, A
31	212	75.4	720	25	US-10-179-131-9251	Sequence 9251, Ap
32	210	74.7	551	19	US-09-538-092-448	Sequence 448, App
33	206	73.3	694	25	US-10-179-131-5254	Sequence 5254, Ap
34	204	72.6	135	21	US-09-791-537-79047	Sequence 79047, A
35	204	72.6	192	21	US-09-417-507-28191	Sequence 28191, A
36	204	72.6	269	21	US-09-791-537-7384	Sequence 7384, Ap
37	204	72.6	269	21	US-09-791-537-80768	Sequence 80768, A
38	204	72.6	298	18	US-09-487-558B-230	Sequence 230, App
39	204	72.6	298	18	US-09-487-558B-230	Sequence 230, App
40	204	72.6	298	22	US-09-801-368-230	Sequence 230, App
41	201	71.5	730	18	US-09-487-558B-126	Sequence 126, App
42	201	71.5	730	18	US-09-487-558B-126	Sequence 126, App
43	201	71.5	730	22	US-09-801-368-126	Sequence 126, App
44	195	69.4	383	21	US-09-791-537-46380	Sequence 46380, A
45	195	69.4	388	21	US-09-791-537-29921	Sequence 29921, A

ALIGNMENTS

RESULT 1  
US-09-791-537-26785

; Sequence 26785, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY-ME  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791.537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 26785  
; LENGTH: 1036  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
US-09-791-537-26785

Query Match 100.0%; Score 281; DB 21; Length 1036;  
Best Local Similarity 100.0%; Pred. No. 3.8e-27;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCTQTTPPLWRRNPDPGQPLCNACGLFLKHLGVVRRPLSLKTDVTKRNR 50  
DB 743 CTNCTQTTPPLWRRNPDPGQPLCNACGLFLKHLGVVRRPLSLKTDVTKRNR 792

## RESULT 2

US-09-791-537-121036  
; Sequence 121036, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 121036  
; LENGTH: 1036  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
US-09-791-537-121036

Query Match 100.0%; Score 281; DB 21; Length 1036;  
Best Local Similarity 100.0%; Pred. No. 3.8e-27;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 50  
|||||  
DB 743 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 792  
|||||

## RESULT 3

US-09-791-537-44327  
; Sequence 44327, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 44327  
; LENGTH: 66  
; TYPE: PRT  
; ORGANISM: pdb 4GATA  
US-09-791-537-44327

Query Match 98.6%; Score 277; DB 21; Length 66;  
Best Local Similarity 98.0%; Pred. No. 6.3e-28;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 50  
|||||  
DB 12 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 61  
|||||

## RESULT 4

US-09-791-537-118817  
; Sequence 118817, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 118817

; LENGTH: 725

; TYPE: PRT  
; ORGANISM: Penicillium chrysogenum  
US-09-791-537-118817

Query Match 98.6%; Score 277; DB 21; Length 725;  
Best Local Similarity 98.0%; Pred. No. 8.6e-27;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 50  
|||||  
DB 525 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 574  
|||||

## RESULT 5

US-09-791-537-125732  
; Sequence 125732, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 125732  
; LENGTH: 860  
; TYPE: PRT  
; ORGANISM: Penicillium roqueforti  
US-09-791-537-125732

Query Match 98.6%; Score 277; DB 21; Length 860;  
Best Local Similarity 98.0%; Pred. No. 1e-26;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 50  
|||||  
DB 660 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 709  
|||||

## RESULT 6

US-09-791-537-116651  
; Sequence 116651, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 116651  
; LENGTH: 862  
; TYPE: PRT  
; ORGANISM: Penicillium chrysogenum  
US-09-791-537-116651

Query Match 98.6%; Score 277; DB 21; Length 862;  
Best Local Similarity 98.0%; Pred. No. 1e-26;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 50  
|||||  
DB 662 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGHWVRPLSLKTDVIKKNR 711  
|||||

## RESULT 7

US-09-791-537-118818  
; Sequence 118818, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 118818  
; LENGTH: 865  
; TYPE: PRT  
; ORGANISM: Penicillium urticae  
US-09-791-537-118818  
Query Match 98.6%; Score 277; DB 21; Length 865;  
Best Local Similarity 98.0%; Pred. No. 1e-26; Mismatches 1; Indels 0; Gaps 0;  
Matches 49; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CTNCFTQTTPLWRRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKRRN 50  
|||||  
Db 665 CTNCFTQTTPLWRRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKRRN 714  
|||||  
RESULT 8  
US-08-750-458-2  
; Sequence 2, Application US/08750458  
; GENERAL INFORMATION:  
; APPLICANT: Christensen, Tove  
; APPLICANT: Hynes, Michael J.  
; TITLE OF INVENTION: A Fungus Wherein The Area Gene Has Been Modified And An Area  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Novo Nordisk of North America, Inc.  
; STREET: 405 Lexington Avenue, Suite 6400  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/750,458  
; FILING DATE: 3-December-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rozek, Carol E.  
; REGISTRATION NUMBER: 36,993  
; REFERENCE/DOCKET NUMBER: 4129.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 866 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-750-458-2  
Query Match 98.6%; Score 277; DB 11; Length 866;  
Best Local Similarity 98.0%; Pred. No. 1e-26; Mismatches 1; Indels 0; Gaps 0;  
Matches 49; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CTNCFTQTTPLWRRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKRRN 50  
|||||

Db 664 CTNCFTQTTPLWRRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKRRN 713  
|||||  
RESULT 9  
US-09-791-537-44920  
; Sequence 44920, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 44920  
; LENGTH: 866  
; TYPE: PRT  
; ORGANISM: Aspergillus parasiticus  
US-09-791-537-44920  
Query Match 98.6%; Score 277; DB 21; Length 866;  
Best Local Similarity 98.0%; Pred. No. 1e-26; Mismatches 1; Indels 0; Gaps 0;  
Matches 49; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CTNCFTQTTPLWRRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKRRN 50  
|||||  
Db 664 CTNCFTQTTPLWRRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKRRN 713  
|||||  
RESULT 10  
US-09-791-537-81933  
; Sequence 81933, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 81933  
; LENGTH: 866  
; TYPE: PRT  
; ORGANISM: Aspergillus oryzae  
US-09-791-537-81933  
Query Match 98.6%; Score 277; DB 21; Length 866;  
Best Local Similarity 98.0%; Pred. No. 1e-26; Mismatches 1; Indels 0; Gaps 0;  
Matches 49; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CTNCFTQTTPLWRRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKRRN 50  
|||||  
Db 664 CTNCFTQTTPLWRRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKRRN 713  
|||||  
RESULT 11  
US-09-487-558-44  
; Sequence 44, Application US/09487558  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: Norman, Thea

APPLICANT: Royer, John  
APPLICANT: Salama, Sofie  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeff  
APPLICANT: Summers, Eric  
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production  
FILE REFERENCE: 109272.147  
CURRENT APPLICATION NUMBER: US/09/487,558  
PRIOR APPLICATION NUMBER: US/09/801,368  
PRIOR FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: US 09/487,558  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: US 60/160,587  
PRIOR FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 440  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 44  
LENGTH: 876  
TYPE: PRT  
ORGANISM: Aspergillus nidulans  
US-09-487-558-44

Query Match 98.6%; Score 277; DB 18; Length 876;  
Best Local Similarity 98.0%; Pred. No. 1.1e-26;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 50  
|||||  
Db 673 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 722

## RESULT 12

US-09-487-558B-44

Sequence 44, Application US/09487558B  
GENERAL INFORMATION:  
APPLICANT: Busby, Robert  
APPLICANT: Cali, Brian  
APPLICANT: Hecht, Peter  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin  
APPLICANT: Maxon, Mary  
APPLICANT: Milne, Todd  
APPLICANT: Norman, Thea  
APPLICANT: Royer, John  
APPLICANT: Salama, Sofie  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeff  
APPLICANT: Summers, Eric  
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
FILE REFERENCE: 109272.130  
CURRENT APPLICATION NUMBER: US/09/487,558B  
CURRENT FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: US 60/487,558  
PRIOR FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 446  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 44  
LENGTH: 876  
TYPE: PRT  
ORGANISM: Aspergillus nidulans  
US-09-487-558B-44

Query Match 98.6%; Score 277; DB 18; Length 876;  
Best Local Similarity 98.0%; Pred. No. 1.1e-26;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 50  
|||||  
Db 673 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 722

RESULT 13  
US-09-791-537-17502  
Sequence 44, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Bionomix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
TITLE OF INVENTION: METHODS OF USE THEREOF  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 17502  
LENGTH: 876  
TYPE: PRT  
ORGANISM: Emmericella nidulans  
US-09-791-537-17502

Query Match 98.6%; Score 277; DB 21; Length 876;  
Best Local Similarity 98.0%; Pred. No. 1.1e-26;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 50  
|||||  
Db 673 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 722

## RESULT 14

US-09-791-537-47207

Sequence 47207, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Bionomix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
TITLE OF INVENTION: METHODS OF USE THEREOF  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 47207  
LENGTH: 876  
TYPE: PRT  
ORGANISM: Emmericella nidulans  
US-09-791-537-47207

Query Match 98.6%; Score 277; DB 21; Length 876;  
Best Local Similarity 98.0%; Pred. No. 1.1e-26;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 50  
|||||  
Db 673 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 722

## RESULT 15

US-09-801-368-44

Sequence 44, Application US/09801368  
GENERAL INFORMATION:  
APPLICANT: Busby, Robert  
APPLICANT: Cali, Brian  
APPLICANT: Hecht, Peter  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin  
APPLICANT: Maxon, Mary  
APPLICANT: Milne, Todd  
APPLICANT: Norman, Thea  
APPLICANT: Royer, John  
APPLICANT: Salama, Sofie  
APPLICANT: Sherman, Amir

Query Match 98.6%; Score 277; DB 18; Length 876;  
Best Local Similarity 98.0%; Pred. No. 1.1e-26;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 50  
|||||  
Db 673 CTNCFTQTTPLRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 722

```

; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 876
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; US-09-801-368-44

```

```

Query Match          98.6%; Score 277; DB 22; Length 876;
Best Local Similarity 98.0%; Pred. No. 1.1e-26;
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKKRNR 50
        |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      673 CTNCFQTTPPLWRNPDGQPLCNACGLFLKHLHGVRPLSLKTDVIKKRNR 722

```

Search completed: January 3, 2003, 19:59:15  
Job time : 188.5 secs







Query Match 91.1%; Score 256; DB 5; Length 50;  
Best Local Similarity 92.0%; Pred. No. 1.5e-24;  
Matches 46; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTNCFQTQTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 50  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 1 CTNCGTKTTPWRRNPQGPQPLCNACGLFLKLGHVVRPLSLKTDVIKKRQR 50  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

## RESULT 3

US-09-724-676-87510  
; Sequence 87510, Application US/09724676

; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 87510  
; LENGTH: 347  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-87510

Query Match 66.9%; Score 188; DB 5; Length 347;  
Best Local Similarity 62.0%; Pred. No. 2.1e-15;  
Matches 31; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 1 CTNCFQTQTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 50  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 196 CANCHTTTTTLWRRNNAEGPVCNACGLYMKLHGVPRLPAMKKEGIQTRKR 245  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

## RESULT 4

US-09-724-676A-87510  
; Sequence 87510, Application US/09724676A

; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676A  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 87510  
; LENGTH: 347  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676A-87510

Query Match 66.9%; Score 188; DB 5; Length 347;  
Best Local Similarity 62.0%; Pred. No. 2.1e-15;  
Matches 31; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 1 CTNCFQTQTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 50  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 196 CANCHTTTTTLWRRNNAEGPVCNACGLYMKLHGVPRLPAMKKEGIQTRKR 245  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

## RESULT 5

US-09-724-676-92485  
; Sequence 92485, Application US/09724676

; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 92485

; LENGTH: 357

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-92485

Query Match 64.8%; Score 182; DB 5; Length 357;  
Best Local Similarity 64.0%; Pred. No. 1.2e-14;  
Matches 32; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 CTNCFQTQTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 50  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 258 CTNCGTTTTTLWRRNAGDPVCNACGLYKLGHVNRPLTMRKDGIQTRNR 307  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

## RESULT 6

US-09-724-676A-92485  
; Sequence 92485, Application US/09724676A

; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676A  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 92485  
; LENGTH: 357  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676A-92485

Query Match 64.8%; Score 182; DB 5; Length 357;  
Best Local Similarity 64.0%; Pred. No. 1.2e-14;  
Matches 32; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 CTNCFQTQTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 50  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 258 CTNCGTTTTTLWRRNAGDPVCNACGLYKLGHVNRPLTMRKDGIQTRNR 307  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

## RESULT 7

US-09-724-676-92294  
; Sequence 92294, Application US/09724676

; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen  
; CURRENT APPLICATION NUMBER: US/09/724,676  
; CURRENT FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 97222  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 92294  
; LENGTH: 455  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-724-676-92294

Query Match 63.3%; Score 178; DB 5; Length 455;  
Best Local Similarity 60.0%; Pred. No. 4.5e-14;  
Matches 30; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 CTNCFQTQTPLWRRNPDGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRNR 50  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
Db 317 CANCHTTTTTLWRRNAGDPVCNACGLYKLGHVNRPLTMRKKEGIQTRNR 366  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

## RESULT 8

US-09-724-676A-92294  
; Sequence 92294, Application US/09724676A

; GENERAL INFORMATION:  
; APPLICANT: Compugen LTD  
; TITLE OF INVENTION: Variants of alternative splicing  
; FILE REFERENCE: 129181.4 Compugen

RESULT 11  
US-09-724-676-87509



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:51:53 ; Search time 21.5 seconds  
(without alignments)  
44.073 Million cell updates/sec

Title: US-09-725-010-3

Perfect score: 281

Sequence: 1 CTNCFQTTPRLRRNPDPG.....LHGVRPLSLKTDVIKRRNR 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US03\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	277	98.6	876	10	US-09-801-368-44
2	204	72.6	298	10	US-09-801-368-230
3	201	71.5	730	10	US-09-801-368-126
4	183	65.1	442	10	US-09-749-728B-11
5	179	63.7	532	10	US-09-801-368-360
6	178	63.3	549	10	US-09-801-368-358
7	158	56.2	128	10	US-09-925-300-1573
8	93	33.1	359	9	US-10-029-180-68
9	93	33.1	440	9	US-10-029-180-124
10	93	33.1	451	9	US-10-029-180-123
11	75	26.7	588	10	US-09-801-368-46
12	56.5	20.1	136	10	US-09-948-018-6
13	56.5	20.1	355	10	US-09-948-018-2
14	56.5	20.1	380	10	US-09-948-018-36
15	56.5	20.1	1194	9	US-09-738-626-5832
16	54	19.2	74	10	US-09-764-846-181
17	53	18.9	108	10	US-09-799-946-11
18	53	18.9	387	10	US-09-967-552A-36
19	52.5	18.7	2237	12	US-10-033-026-8

#### ALIGNMENTS

##### RESULT 1

US-09-801-368-44

; Sequence 44, Application US/09801368

; Patent No. US20020128250A1

; GENERAL INFORMATION:

; APPLICANT: Busby, Robert

; APPLICANT: Cali, Brian

; APPLICANT: Hecht, Peter

; APPLICANT: Holtzman, Doug

; APPLICANT: Madden, Kevin

; APPLICANT: Maxon, Mary

; APPLICANT: Milne, Todd

; APPLICANT: No. US20020128250Alman, Thea

; APPLICANT: Royer, John

; APPLICANT: Salama, Sofie

; APPLICANT: Sherman, Amir

; APPLICANT: Silva, Jeff

; APPLICANT: Summers, Eric

; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

; FILE REFERENCE: 109272.147

; CURRENT APPLICATION NUMBER: US/09/801,368

; CURRENT FILING DATE: 2001-03-07

; PRIOR APPLICATION NUMBER: US 09/487,558

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: US 60/160,587

; PRIOR FILING DATE: 1999-10-20

; NUMBER OF SEQ ID NOS: 440

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 44

; LENGTH: 876

; TYPE: PRT

; ORGANISM: Aspergillus nidulans

US-09-801-368-44

Query Match 98.6%; Score 277; DB 10; Length 876;

Best Local Similarity 98.0%; Pred. No. 7.9e-29;

Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPRLRRNPDPGCPICNACGLFLKLGVRPLSLKTDVIKRRNR 50

|||||

DB 673 CTNCFQTTPRLRRNPDPGCPICNACGLFLKLGVRPLSLKTDVIKRRNR 722



;; CURRENT APPLICATION NUMBER: US/09/801.368  
;; CURRENT FILING DATE: 2001-03-07  
;; PRIOR APPLICATION NUMBER: US 09/487,558  
;; PRIOR FILING DATE: 2000-01-19  
;; PRIOR APPLICATION NUMBER: US 60/160,587  
;; PRIOR FILING DATE: 1999-10-20  
;; NUMBER OF SEQ ID NOS: 440  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 360  
;; LENGTH: 532  
;; TYPE: PRT  
;; ORGANISM: Penicillium chrysogenum  
US-09-801-368-360

Query Match 63.7%; Score 179; DB 10; Length 532;  
Best Local Similarity 58.0%; Pred. No. 4.4e-16;  
Matches 29; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

Qy 1 CTNCFQTPTPLWRRNPDGQPLCNACGLFLKLHGYYVVRPLSLKTDVKKRNR 50  
||| ||||| :|:|||||: |||| || :||:|  
Db 238 CQCGTTVPLWRRDQGHPCNACGLYKLGCGYRPTNMKKSIIKKRR 287

## RESULT 6

US-09-801-368-358  
;; Sequence 358, Application US/09801368  
;; Patent No. US20020128250A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Busby, Robert  
;; APPLICANT: Cali, Brian  
;; APPLICANT: Hecht, Peter  
;; APPLICANT: Holtzman, Doug  
;; APPLICANT: Madden, Kevin  
;; APPLICANT: Maxon, Mary  
;; APPLICANT: Milne, Todd  
;; APPLICANT: No. US20020128250A1man, Thea  
;; APPLICANT: Royer, John  
;; APPLICANT: Salama, Sofie  
;; APPLICANT: Sherman, Amir  
;; APPLICANT: Silva, Jeff  
;; APPLICANT: Summers, Eric  
;; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
;; FILE REFERENCE: 109272.147  
;; CURRENT APPLICATION NUMBER: US/09/801.368  
;; CURRENT FILING DATE: 2001-03-07  
;; PRIOR APPLICATION NUMBER: US 09/487,558  
;; PRIOR FILING DATE: 2000-01-19  
;; PRIOR APPLICATION NUMBER: US 60/160,587  
;; PRIOR FILING DATE: 1999-10-20  
;; NUMBER OF SEQ ID NOS: 440  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 358  
;; LENGTH: 549  
;; TYPE: PRT  
;; ORGANISM: Aspergillus nidulans  
US-09-801-368-358

Query Match 63.3%; Score 178; DB 10; Length 549;  
Best Local Similarity 58.0%; Pred. No. 6.2e-16;  
Matches 29; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

Qy 1 CTNCFQTPTPLWRRNPDGQPLCNACGLFLKLHGYYVVRPLSLKTDVKKRNR 50  
||| ||||| :|:|||||: |||| || :||:|  
Db 251 CQCGTTVPLWRRDQGHPCNACGLYKLGCGYRPTNMKKSIIKKRR 300

## RESULT 7

US-09-925-300-1573  
;; Sequence 1573, Application US/09925300  
;; Patent No. US20020151681A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Craig Rosen,  
;; APPLICANT: Steve Ruben

;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
;; FILE REFERENCE: PA101  
;; CURRENT APPLICATION NUMBER: US/09/925,300  
;; CURRENT FILING DATE: 2001-08-10  
;; PRIOR APPLICATION NUMBER: PCI/US00/05988  
;; PRIOR FILING DATE: 2000-03-08  
;; PRIOR APPLICATION NUMBER: 60/124,270  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 1890  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 1573  
;; LENGTH: 128  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-925-300-1573

Query Match 56.2%; Score 158; DB 10; Length 128;  
Best Local Similarity 62.2%; Pred. No. 5.3e-14;  
Matches 28; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Qy 6 TOTTPLWRRNPDGQPLCNACGLFLKLHGYYVVRPLSLKTDVKKRNR 50  
||| ||||| :|:|||||: |||| || :||:|  
Db 2 TTTTLWRRNANGPVCNACGLYKLGHNVRPLTMKKGIOQRNR 46

## RESULT 8

US-10-029-180-68  
;; Sequence 68, Application US/10029180  
;; Publication No. US20020182708A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Cali, Brian M.  
;; APPLICANT: Holtzman, Doug  
;; APPLICANT: Madden, Kevin T.  
;; APPLICANT: Milna, G. Todd  
;; APPLICANT: Sherman, Amir  
;; APPLICANT: Silva, Jeffery C.  
;; APPLICANT: Trueheart, Josh  
;; APPLICANT: Zhang, Lixin  
;; TITLE OF INVENTION: No. US20020182708A1el Regulators of Fungal Gene Expression  
;; FILE REFERENCE: MIC-004  
;; CURRENT APPLICATION NUMBER: US/10/029,180  
;; CURRENT FILING DATE: 2001-12-22  
;; PRIOR APPLICATION NUMBER: US 60/257,431  
;; PRIOR FILING DATE: 2000-12-22  
;; NUMBER OF SEQ ID NOS: 138  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 68  
;; LENGTH: 359  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: fungal gene  
US-10-029-180-68

Query Match 33.1%; Score 93; DB 9; Length 359;  
Best Local Similarity 57.6%; Pred. No. 7.2e-05;  
Matches 19; Conservative 3; Mismatches 9; Indels 2; Gaps 2;

Qy 1 CTNCFQTPTPLWRRNPDG-QPLCNACGL-FLKL 31  
|:| ||||| :|:|||||: |||| || :||:|  
Db 308 CHCNSRSETPWRRGPDGPRTLCNACGLHYAKL 340

## RESULT 9

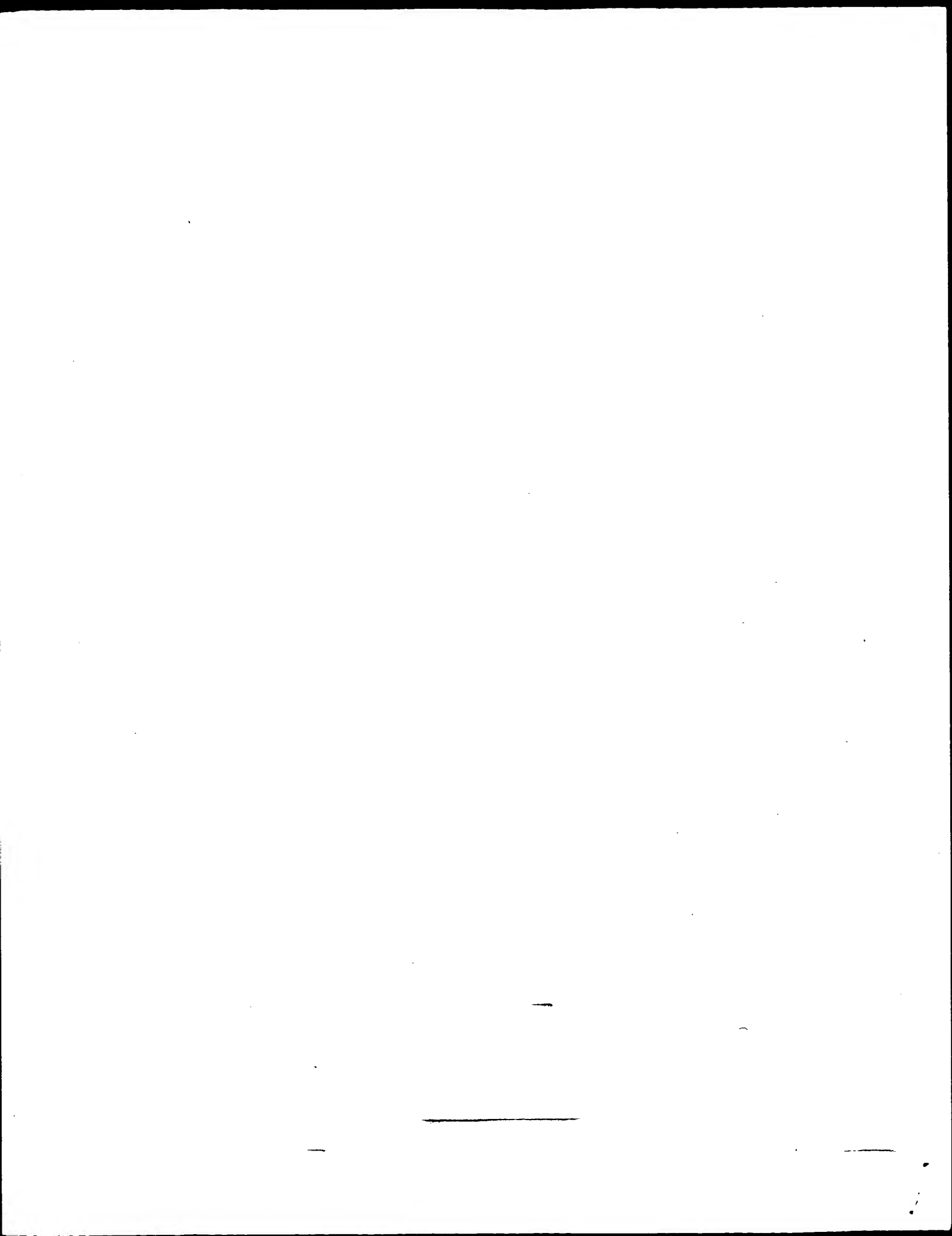
US-10-029-180-124  
;; Sequence 124, Application US/10029180  
;; Publication No. US20020182708A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Cali, Brian M.  
;; APPLICANT: Holtzman, Doug  
;; APPLICANT: Madden, Kevin T.  
;; APPLICANT: Milna, G. Todd  
;; APPLICANT: Sherman, Amir

US 09 348 010 2  
; Sequence 2, Application US/09948018  
; Patent No. US20020150977A1



RESULT 15  
 US-09-738-626-5832  
 ; Sequence 5832, Application US/09738626  
 ; Publication No. US20020197605A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAGAWA, SATOSHI  
 ; APPLICANT: MIZOGUCHI, HIROSHI  
 ; APPLICANT: ANDO, SEIKO  
 ; APPLICANT: HAYASHI, MIKIRO  
 ; APPLICANT: OCHIAI, KEIRO  
 ; APPLICANT: YOKOI, HARUHIKO  
 ; APPLICANT: TATEISHI, NAKKO  
 ; APPLICANT: SENOH, AKIHIRO  
 ; APPLICANT: IKEDA, MASATO  
 ; APPLICANT: OZAKI, AKIO  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-125  
 ; CURRENT APPLICATION NUMBER: US/09/738,626  
 ; CURRENT FILING DATE: 2000-12-18

Search completed: January 3, 2003, 20:01:18  
Job time : 21.5 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 18:48:02 ; Search time 44.5 Seconds  
(without alignments)  
149.720 Million cell updates/sec

Title: US-09-725-010-3

Perfect score: 281

Sequence: 1 CTCNCTQTTLPLWRNPDGQP.....LHGVRPLSLKTDVIKKNR 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	281	100.0	50	22	AA882413
2	277	98.6	853	20	AAW81079
3	277	98.6	866	17	AA88408
4	277	98.6	866	18	AAW31630
5	256	91.1	50	22	AA882414
6	190	67.6	950	22	ABB71271
7	188	66.9	279	22	AA874577
8	188	66.9	279	23	ABP41885
9	188	66.9	449	19	AAW56703
10	183	65.1	442	22	AA899920

11	183	65.1	442	22	AA864319	Human GATA4 protei
12	183	65.1	442	22	AA864849	Heart muscle cell
13	180	64.1	486	22	ABB64550	Drosophila melanog
14	178	63.3	443	23	AAU97800	Mouse GATA-3 trans
15	178	63.3	443	23	AAU97801	Transgenic animal
16	178	63.3	443	23	ABB85021	Pain regulated pro
17	178	63.3	443	23	ABB85022	Pain regulated pro
18	178	63.3	444	20	AA423279	GATA-3 amino acid
19	176	62.6	734	22	ABB71878	Drosophila melanog
20	158	56.2	128	21	AA856995	Human prostate can
21	139.5	49.6	785	22	AA892512	Human protein sequ
22	139.5	49.6	1281	21	AA800190	Breast cancer prot
23	139.5	49.6	1281	22	AAE02189	Human breast cance
24	128.5	45.7	1294	21	AA840550	Human ORFX ORF314
25	117	41.6	273	22	AA878812	Zinc finger protei
26	108.5	38.6	187	21	AA815686	Arabidopsis thalia
27	108.5	38.6	271	21	AA815685	Arabidopsis thalia
28	108.5	38.6	309	21	AA815684	Arabidopsis thalia
29	108	38.4	119	22	AA825514	Human protein sequ
30	107	38.1	210	21	AA841113	Human ORFX ORF877
31	100.5	35.8	276	21	AA808697	Arabidopsis thalia
32	100.5	35.8	276	21	AA842639	Arabidopsis thalia
33	100.5	35.8	294	21	AA808696	Arabidopsis thalia
34	100.5	35.8	294	21	AA842638	Arabidopsis thalia
35	100.5	35.8	295	21	AA808695	Arabidopsis thalia
36	100.5	35.8	295	21	AA842637	Arabidopsis thalia
37	100	35.6	163	21	AA833111	Pinus radiata tran
38	97	34.5	264	23	AAU93001	Arabidopsis trans
39	96.5	34.3	278	21	AA811326	Arabidopsis thalia
40	96.5	34.3	285	21	AA821607	Arabidopsis thalia
41	96.5	34.3	287	21	AA811325	Arabidopsis thalia
42	96.5	34.3	295	21	AA821606	Arabidopsis thalia
43	96.5	34.3	297	21	AA811324	Arabidopsis thalia
44	96.5	34.3	329	21	AA821605	Arabidopsis thalia
45	96	34.2	240	23	AAU93121	Arabidopsis trans

#### ALIGNMENTS

RESULT 1  
AA882413  
ID AA882413 standard; Protein; 50 AA.  
XX  
AC AA882413;  
XX  
DT 06-AUG-2001 (first entry)  
DE NIT2 DNA binding domain.  
DE  
DE  
XX  
KW Hyphal wall protein; Hwp1; hyphal-specific gene; infection;  
KW candidiasis; therapy; antifungal; fungicide; transcription factor;  
KW NIT2; DNA binding protein.  
XX  
OS Neurospora crassa.  
XX  
PN WO200138550-A2.  
XX  
PD 31-MAY-2001.  
XX  
PF 29-NOV-2000; 2000WO-US32464.  
XX  
PR 29-NOV-1999; 99US-0167672.  
XX  
PA (SUND/) SUNDSTROM P.  
XX  
PI Sundstrom P;  
XX  
DR WPI; 2001-367698/38.  
XX  
PT Interfering with expression of hyphal-specific genes in fungus for  
PT inhibiting fungal cell growth involves interfering with transcription  
PT of hyphal-specific genes mediated by cis acting sequences

XX SQ Sequence 50 AA;

Query Match 100.0%; Score 281; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-29;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50  
 |||||  
 DB 1 CTNCFQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50

RESULT 2  
 ID AAW81079  
 AC AAW81079;  
 DT 24-MAY-1999 (first entry)  
 DE Amino acid sequence of the Aspergillus oryzae area protein.  
 KW area gene; peptidase; Koji mold; L-glutamine; proliase;  
 KW prolyl-dipeptidyl-peptidase.  
 XX Aspergillus oryzae.  
 XX WO9902691-A1.  
 XX 21-JAN-1999.  
 XX 01-MAY-1998; 98WO-EF02785.  
 XX 05-JUL-1997; 97EP-0111378.  
 XX (NEST ) SOC PROD NESTLE SA.  
 XX Affolter M, Van Den Broek P;  
 XX WPI; 1999-120892/10.  
 XX N-PSDB; AAV68650.  
 XX New koji mold capable of expressing two times more endo- and  
 PT exo-peptidases than wild-type strain Aspergillus oryzae CNCM I-1882

Claim 31; Page 80; 95pp; English.

The present sequence is that of the NIT2 DNA binding domain of *Neurospora crassa*. The NIT2 protein is a global positive-acting transcription factor of nitrogen structural genes when preferred N-sources are lacking. The promoter region (see AAF90471) of the *Candida albicans* hyphal-specific HWPI gene contains numerous binding sites for NIT2. GAT99 (see AAB82414), showing 92% homology to NIT2, was identified in *C. albicans*, suggesting that regulation of HWPI expression may be tied to nitrogen regulatory events. The HWPI gene encodes an adhesin required for the pathogenesis of candidiasis. The invention provides a method for interfering with the expression of hyphal-specific genes in a fungus resulting in inhibition of cell growth. This involves interfering with the transcription of the hyphal-specific gene mediated by cis-acting sequences. The interfering step may involve manipulating the binding of a DNA binding protein to a cis-regulatory element. The fungus may be a non-pathogenic or a pathogenic fungus, such as *C. albicans*, and the hyphal-specific gene may be HWPI, in which the NIT2 binding sites act as the cis-regulatory elements. The method provides a means of treating infection of mammalian hosts such as immunocompromised or immunosuppressed humans, including those having AIDS or undergoing transplantation or anti-cancer therapy, burns patients, patients with diabetic ketoacidosis, and patients in which the normal microbial flora has been disrupted because of disease, trauma or chemical, radiation or other immunosuppressive prophylaxis. Also provided is a method for characterising genes under control of a DNA binding protein.

PT - useful for hydrolyzing protein-containing materials, and producing food products with improved organoleptic properties

XX Claim 11; Pages 28-30; 53pp; English.

XX This is the amino acid sequence of the area protein of *Aspergillus oryzae* used in the method of the invention involving the doubled expression of peptidases in a new Koji mold. The koji mold is useful for hydrolyzing protein-containing materials, and for over-producing proteolytic enzymes. When hydrolyzing protein-containing material (containing 5 mM L-glutamine), the mold may be used in combination with an enzyme and/or a microorganism providing a prolidase activity. Additionally, the mold is useful in fermentation processes for creating food products. The koji mold is used to produce high levels of endopeptidases and exopeptidases, including enhanced prolyl-dipeptidyl-peptidase activity, which may be useful for releasing L-glutamine from peptides.

XX SQ Sequence 853 AA;

Query Match 98.6%; Score 277; DB 20; Length 853;  
 Best Local Similarity 98.0%; Pred. No. 1.3e-27;  
 Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50  
 |||||  
 DB 652 CTNCFQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 701

RESULT 3  
 ID AAR88408  
 AC AAR88408;  
 DT 26-JUN-1996 (first entry)  
 DE Area activator protein.  
 KW Area activator; protease-free; host cell.  
 XX *Aspergillus oryzae* strain IFO4177.  
 XX WO9535385-A1.  
 XX 28-DEC-1995.  
 XX 19-JUN-1995; 95WO-DK00254.  
 XX 17-JUN-1994; 94DK-0000717.  
 XX (NOVO ) NOVO-NORDISK AS.  
 XX Christensen T, Hynes MJ;  
 XX WPI; 1996-058422/06.  
 XX N-PSDB; AAT09946.  
 XX Fungus with a modified area gene - useful to produce proteins susceptible to degradation by protease(s)  
 XX Claim 23; Page 23-25; 42pp; English.  
 XX The Area activator protein (AAR88408) of *Aspergillus oryzae* IFO4177 controls nitrogen catabolism and influences the production of extracellular proteases. It is the product of the area gene (AAT09946). Inactivation of the activator, by deletion or antisense inhibition of the gene, yields protease-free *Aspergillus* cells, e.g. strain ToC913, suitable as hosts for prodn. of protease-susceptible heterologous proteins such as enzymes and industrial or therapeutic proteins.

QY 1 CTNCFTQTTPWRRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 50  
|||||:|||||  
664 CTNCFTQTTPWRRNPDGQPLCNACGLFLKLHGVRPLSLKTDVIKKNR 713

QY 1 CTNCFTQTTPWRRRNPDPGLCNACGLFLKLHGVRPLSLKTDVKKRNR 50  
|||||:|||||





XX Differentiation; heart muscle cell; cytokine; transcription factor;  
 KW proliferation; surface antigen; heart disease; cardiomyocyte;  
 KW bone marrow; umbilical blood cell; heart muscle degeneration;  
 KW myocardial infarction.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200148150-A1.  
 XX  
 PD 05-JUL-2001.  
 XX  
 PF 02-NOV-2000; 2000WO-JP07741.  
 XX  
 PR 28-DEC-1999; 99JP-0372826.  
 XX  
 PR 28-FEB-2000; 2000WO-JP01148.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX  
 XX Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K, Gojo S;  
 PI Yamada Y;  
 XX WPI; 2001-425655/45.  
 DR N-PSDB; AAH44356.  
 XX  
 XX Cells capable of differentiating into cardiomyocytes and originating in  
 PT bone marrow or umbilical blood cells for study of cardiomyocyte  
 PT differentiation and treatment of heart disease -  
 XX  
 PS Claim 41; Page 98-99; 187pp; Japanese.  
 XX  
 CC The present invention describes cells originating in bone marrow or  
 CC umbilical blood cells which are capable of differentiating into  
 CC cardiomyocytes. Also described are: (1) cardiomyocytes produced by the  
 CC differentiation of the cells; (2) a method for carrying out the  
 CC differentiation into cardiomyocytes, regulated by a promotional and/or  
 CC inhibitory factor; (3) a method for the differentiation of the cells  
 CC into cell types other than cardiomyocytes; (4) drug compositions  
 CC promoting the formation of heart muscle and regeneration of heart tissue  
 CC which contain the cells; (5) a method for the production of antibodies  
 CC which recognise the cells, especially antibodies which recognise a  
 CC surface antigen on the cells; (6) a method for screening factors which  
 CC promote the proliferation of the cells; (7) a method for immortalising  
 CC the cells by expressing telomerase in them; (8) drug compositions for  
 CC the treatment of heart disease which contain the immortalised cells; and  
 CC (9) cell-free supernatant from the culture of the cells and its use in  
 CC promoting their differentiation into cardiomyocytes. The cells are used  
 CC in the treatment of diseases involving heart muscle degeneration, such  
 CC as myocardial infarction and in the study of cardiomyocyte  
 CC differentiation. AAH44351 to AAH44409 and AAB99915 to AAB99935 represent  
 CC sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 442 AA;  
 Query Match 65.1%; Score 183; DB 22; Length 442;  
 Best Local Similarity 60.0%; Pred. No. 1.4e-15;  
 Matches 30; Conservative 10; Mismatches 10; Indels 0; Gaps 0;  
 QY 1 CTNCFQTTPPLWRNPDGQPLNACGLFLKLGWVRPLSLKTDVTKRNR 50  
 DB 271 CANGCQTTTTLWRNABGEPCVNCACGLYMKLHGVRPLAMRKEGIQTRKR 320  
 RESULT 11  
 AAG64319  
 ID AAG64319 standard; Protein; 442 AA.  
 XX  
 AC AAG64319;  
 XX  
 XX 24-SEP-2001 (first entry)  
 DT Human GATA4 protein.  
 DE  
 XX

KW Angiogenesis; cardiant; cell differentiating agent; bone marrow;  
 KW heart muscle cell; heart disease; human; GATA4.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200148149-A1.  
 XX  
 PD 05-JUL-2001.  
 XX  
 PF 28-FEB-2000; 2000WO-JP01148.  
 XX  
 PR 28-DEC-1999; 99JP-0372826.  
 XX  
 PR (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX  
 XX Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K;  
 PI WPI; 2001-418252/44.  
 DR N-PSDB; AAH49591.  
 XX  
 XX New adult bone marrow-originated cells capable of differentiating into  
 PT heart muscle cells, applicable as remedies for various heart diseases  
 PT particularly with damaged heart muscle accompanying degeneration -  
 XX  
 XX Claim 23; Pages 74-76; 158pp; Japanese.  
 XX  
 CC The present invention relates to cells isolated from bone marrow, which  
 CC are capable of at least differentiating into heart muscle cells. The  
 CC cells are applicable as remedies for various heart diseases particularly  
 CC with damaged heart muscle accompanying degeneration. The present sequence  
 CC was used to illustrate the present invention.  
 XX  
 SQ Sequence 442 AA;  
 Query Match 65.1%; Score 183; DB 22; Length 442;  
 Best Local Similarity 60.0%; Pred. No. 1.4e-15;  
 Matches 30; Conservative 10; Mismatches 10; Indels 0; Gaps 0;  
 QY 1 CTNCFQTTPPLWRNPDGQPLNACGLFLKLGWVRPLSLKTDVTKRNR 50  
 DB 271 CANGCQTTTTLWRNABGEPCVNCACGLYMKLHGVRPLAMRKEGIQTRKR 320  
 RESULT 12  
 AAG64849  
 ID AAG64849 standard; Protein; 442 AA.  
 XX  
 AC AAG64849;  
 XX  
 XX 21-SEP-2001 (first entry)  
 DT Heart muscle cell differentiation related protein SEQ ID NO: 11.  
 DE Heart muscle cell; human; cell differentiation; heart disease.  
 XX  
 KW Heart muscle cell; human; cell differentiation; heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200148151-A1.  
 XX  
 PD 05-JUL-2001.  
 XX  
 PF 27-DEC-2000; 2000WO-JP09323.  
 XX  
 PR 28-DEC-1999; 99JP-0372826.  
 XX  
 PR 28-FEB-2000; 2000WO-JP01148.  
 XX  
 PR 02-NOV-2000; 2000WO-JP07741.  
 XX  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Umezawa A, Hata J, Fukuda K, Ogawa S, Sakurada K, Gojo S;  
 PI Yamada Y;  
 XX WPI; 2001-425656/45.  
 DR



CC The sequence data for this patent did not form part of the printed  
CC (AB557737) AB5572072).  
CC  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO

XX	
DT	13-AUG-2002 (first entry)
DE	Transgenic animal expressing GATA-3 associated protein.
XX	
KW	Transgenic animal; transcription factor; GATA-3; allergic disease;
KX	autoimmune disease.
XX	
OS	Mus sp.
XX	
PN	JP2002112666-A.
XX	
PD	16-APR-2002.
XX	
Pf	04-OCT-2000; 2000JP-0305253.
XX	
PR	04-OCT-2000; 2000JP-0305253.
XX	
PA	(SUMU ) SUMITOMO SEIYAKU KK.
XX	
DR	WPI; 2002-447972/48.
DR	N-PSDB; ABK51619.
XX	
PT	Transgenic non-human mammals expressing GATA-3 or its variant prepared
PT	by mating a mouse with introduced a transfer factor GATA-3 gene and a
PT	mouse with introduced egg white albumin recognising T-cell receptor
PT	gene -
XX	
PS	Disclosure; Page 6-7; 9pp; Japanese.
XX	
CC	The invention describes a transgenic non-human mammal expressing
CC	transcription factor GATA-3 or its variant. The transgenic animal is
CC	useful for screening for agents for prevention and/or treatment of
CC	allergic diseases and autoimmune diseases. Prepared transgenic non-human
CC	mammals are useful for screening of treatment agent of allergic diseases
CC	and autoimmune diseases. This is the amino acid sequence of a protein
CC	associated with the transgenic animal expressing GATA-3 described in the
CC	invention.
XX	
SQ	Sequence 443 AA;
	Query Match 63.3%; Score 178; DB 23; Length 443;
	Best Local Similarity 60.0%; Pred. No. 6,1e-15;
	Matches 30; Conservative 8; Mismatches 12; Indels 0; Gaps 0;
QY	1 CTNCFTQTTPLRNPDGCPGNACGLFLKHGVVRPLSLKTDTVTKRRN 50
	:   :           :   :         :   :
Dd	317 CACCQTTTTLWRNAGDPVNCAGLYKKHNIINRLPTMKKEGIOTRRN 360

Search completed: January 3, 2003, 19:47:49  
Job time : 45.5 secs

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic -nucleic search, using sw model  
Run on: January 3, 2003, 15:22:02 ; Search time 2756 seconds  
(without alignments)  
15522.904 Million cell updates/sec  
Title: US-09-725-010-1  
Perfect score: 1470  
Sequence: 1 ggatcttttttttcttcttcc.....cttttgatttgcataatg 1470

-Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: gb.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pin.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1470	100.0	1470	6	AX154641	Sequence
2	505	34.4	2682	8	CAU64206	U64206 Candida alb
3	153	10.4	2188	8	AF001978	AF001978 Candida a
c 4	136.2	9.3	42565	8	CAC35A5	AL033396 C.albican
c 5	91.2	6.2	86827	3	PFMAL3P5	AL034556 Plasmodiu
6	86.8	5.9	104992	2	AC005504	AC005504 Plasmodiu
7	86.8	5.9	169546	2	AC004157	AC004157 Plasmodiu
c 8	75.2	5.1	104014	2	AC116921	AC116921 Dictyoste
9	73.4	5.0	14867	3	AE001398	AE001398 Plasmodiu
10	73.4	5.0	164399	3	PFMAL3P6	298551 Plasmodium
11	72.8	5.0	113880	3	PFMAL3P4	AL008970 Plasmodiu
c 12	70.8	4.8	2009	6	AX457067	AX457067 Sequence
c 13	70.4	4.8	192929	2	AC005505	AC005505 Plasmodiu
14	69.4	4.7	56099	2	AC115598	AC115598 Dictyoste
15	69.2	4.7	253305	3	PFMAL3P7	AL034559 Plasmodiu
c 16	68.2	4.6	115489	2	AC117072	AC117072 Dictyoste
c 17	68	4.6	9810	6	AX345328	AX345328 Sequence
18	67.2	4.6	7347	1	AF211124	AF211124 Carsonell
c 19	67.2	4.6	67970	3	PFMAL1P3	AL031746 Plasmodiu
c 20	67	4.6	130349	9	AC011593	AC011593 Homo sapi
c 21	67	4.6	198431	9	AC109994	AC109994 Homo sapi
c 22	66.8	4.5	97371	8	ATT28A8	AL162691 Arabidops
c 23	66.6	4.5	1141	6	AX083744	AX083744 Sequence
24	66.4	4.5	14635	3	AE001423	AE001423 Plasmodiu
c 25	66.4	4.5	18624	6	AX346604	AX346604 Sequence
26	66.4	4.5	30726	2	AC117269	AC117269 Dictyoste
c 27	66.2	4.5	1192	9	HS323759	AJ323759 Homo sapi
c 28	66.2	4.5	115489	2	AC117072	AE001398 Plasmodiu
29	66.2	4.5	115489	2	AC117072	AC117072 Dictyoste
c 30	66.2	4.5	133148	9	AC009435	AC009435 Homo sapi
c 31	66	4.5	17538	6	AX346059	AX346059 Sequence
c 32	66	4.5	201299	9	AC007158	AC007158 Homo sapi
c 33	65.8	4.5	156060	2	AC004153	AC004153 Plasmodiu
c 34	65.6	4.5	7584	6	AX286935	AX286935 Sequence
c 35	65.4	4.4	22977	9	AC106883	AC106883 Homo sapi
c 36	65.4	4.4	162515	9	AC107032	AC107032 Homo sapi
37	65.2	4.4	23676	2	AC115599	AC115599 Dictyoste
c 38	65.2	4.4	124820	2	AC117073	AC117073 Dictyoste
c 39	65.2	4.4	141017	2	AC116962	AC116962 Dictyoste
c 40	65.2	4.4	165260	9	AC024341	AC024341 Homo sapi
c 41	65	4.4	43993	2	AC116965	AC116965 Dictyoste
42	65	4.4	165059	9	AC106856	AC106856 Homo sapi
c 43	65	4.4	321003	2	PFMAL4P3	AL035476 Plasmodiu
c 44	64.8	4.4	1141	6	AX083744	AX083744 Sequence
c 45	64.8	4.4	5884	6	AX347067	AX347067 Sequence

ALIGNMENTS

RESULT 1  
AX154641  
LOCUS  
SEQUENCE 1 from Patent WO0138550.  
DEFINITION  
ACCESSION AX154641  
VERSION AX154641.1 GI:14536200  
KEYWORDS  
SOURCE  
ORGANISM  
Candida albicans.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Patent: WO 0138550-A 1 31-MAY-2001;  
Methods for altering the expression of hyphal-specific genes  
1 (bases 1 to 1470)  
Sundstrom,P.  
1470 bp  
DNA  
linear  
PAT 22-JUN-2001

Sundstrom, Paula (US)		Location/Qualifiers	
source		1..1470	
		/organism="Candida albicans"	
		/db_xref="taxon:5476"	
BASE COUNT	509 a	260 c	186 g 515 t
ORIGIN			
Query Match	100.0%;	Score 1470;	DB 6; Length 1470;
Best Local Similarity	100.0%;	Pred. No. 2e-225;	
Matches 1470;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	1	GGATCTTCTTTTCCATTCCTTAAACCGATCAAGAAAGAAAGTGGAATAAGCTAT	60
Db	1	GGATCTTCTTTTCCATTCCTTAAACCGATCAAGAAAGAAAGTGGAATAAGCTAT	60
QY	61	GATAAATGTTGATTTTGTGTAATCAATCAACTAAGCAGCTTTCACAGTTAAAAAGTACG	120
Db	61	GATAAATGTTGATTTTGTGTAATCAATCAACTAAGCAGCTTTCACAGTTAAAAAGTACG	120
QY	121	TGTGTTGTTCTCGTCGCTCTAATTTCTGTTGACGAGGATTAATAACAAGAAATAACAG	180
Db	121	TGTGTTGTTCTCGTCGCTCTAATTTCTGTTGACGAGGATTAATAACAAGAAATAACAG	180
QY	181	GAACCCCTCCAAAAAATAATTTTGACCTTACAGCAGATATAATGCGGATAAAGTTCG	240
Db	181	GAACCCCTCCAAAAAATAATTTTGACCTTACAGCAGATATAATGCGGATAAAGTTCG	240
QY	241	CATAATAAAACTCTTTGAAACATACGATATGTTATCTTTTCATAACTGGAATATTTT	300
Db	241	CATAATAAAACTCTTTGAAACATACGATATGTTATCTTTTCATAACTGGAATATTTT	300
QY	301	GCTTTTTTTTAAACATATGAACAATGAAAAAAGAAATGAAAGGTAAGAGTTGCC	360
Db	301	GCTTTTTTTTAAACATATGAACAATGAAAAAAGAAATGAAAGGTAAGAGTTGCC	360
QY	361	TAAACATTTGAAATAATAGGCTTAAGGTTTTTCTGATGCGTTTAACTAAAAGGAAATAA	420
Db	361	TAAACATTTGAAATAATAGGCTTAAGGTTTTTCTGATGCGTTTAACTAAAAGGAAATAA	420
QY	421	CAAAAGTTATTAGCGATACCTGCGTAGGTGTCACAAAATATATTTTGCACGTTAGCT	480
Db	421	CAAAAGTTATTAGCGATACCTGCGTAGGTGTCACAAAATATATATTTTGCACGTTAGCT	480
QY	481	CTATAGAAAATATACAACTAAATCCCTTAAGGAATTTCCCTATATATATAGGAATCC	540
Db	481	CTATAGAAAATATACAACTAAATCCCTTAAGGAATTTCCCTATATATATAGGAATCC	540
QY	541	CTCTCACAGTGAAGTAAATATCCATCTGAAATATACGTCACCACTAAATCCATCAATAAAA	600
Db	541	CTCTCACAGTGAAGTAAATATCCATCTGAAATATACGTCACCACTAAATCCATCAATAAAA	600
QY	601	TAGATTAGTGTATTGTTCTCTTCAGTACAAATTAACATTAATGCAATGCTAGCTTATTG	660
Db	601	TAGATTAGTGTATTGTTCTCTTCAGTACAAATTAACATTAATGCAATGCTAGCTTATTG	660
QY	661	TTCAATAATTAGCCATTTGTCACACCCCTAATTCGAACATTAACCTATATTTTCTT	720
Db	661	TTCAATAATTAGCCATTTGTCACACCCCTAATTCGAACATTAACCTATATTTTCTT	720
QY	721	GTCCCTCTTTGTTTTTCTTAACAAAAATGTTCCAGAAATTTTTTAAAAATATTGAAAA	780
Db	721	GTCCCTCTTTGTTTTTCTTAACAAAAATGTTCCAGAAATTTTTTAAAAATATTGAAAA	780
QY	781	AACACATAACACTTTGAGTATGATAATATCACTATTAATGCTGTTTGAAGAGTAAGAAT	840
Db	781	AACACATAACACTTTGAGTATGATAATATCACTATTAATGCTGTTTGAAGAGTAAGAAT	840
QY	841	CAAAATTTTTTCTAACTGCACTAATGCACTTTACATCAACTGGATGTTATTGTCATCTAC	900
Db	841	CAAAATTTTTTCTAACTGCACTAATGCACTTTACATCAACTGGATGTTATTGTCATCTAC	900
QY	901	TACTATAAGCTCAAAACAATAATTTCTTCAAAAATGTTATAATTAACAAGTCACTATAAT	960
Db	901	TACTATAAGCTCAAAACAATAATTTCTTCAAAAATGTTATAATTAACAAGTCACTATAAT	960
QY	961	TCCTTTGGATCCAAAACAAGGAATTCGGAATAATTCGACGATAAATTCGACCTCAACAATTC	1020
Db	961	TCCTTTGGATCCAAAACAAGGAATTCGGAATAATTCGACGATAAATTCGACCTCAACAATTC	1020
QY	1021	ATTGTAAGAGGAGAGATTTTGGTAGGCTCATATATCGCTTATTAATGCTCTAAAGTAA	1080
Db	1021	ATTGTAAGAGGAGAGATTTTGGTAGGCTCATATATCGCTTATTAATGCTCTAAAGTAA	1080
QY	1081	TCFAAACAACAACAACCTTTCTAAACCTTAATAATAACCCCTAATGCTCACAACCGG	1140
Db	1081	TCFAAACAACAACAACCTTTCTAAACCTTAATAATAACCCCTAATGCTCACAACCGG	1140
QY	1141	GATAAGTTAGTTAGCCAGCTGTTTTTTTGGCCCTATTTTATGACTACATATTTGTTT	1200
Db	1141	GATAAGTTAGTTAGCCAGCTGTTTTTTTGGCCCTATTTTATGACTACATATTTGTTT	1200
QY	1201	CACCTTTTGTGGAGCTTTAATACCCGTTTTTGCACCTCTCTTTGATCACCTGTATCCG	1260
Db	1201	CACCTTTTGTGGAGCTTTAATACCCGTTTTTGCACCTCTCTTTGATCACCTGTATCCG	1260
QY	1261	CTTTTAAACATAGCAACTCTTTGTAAGTCCCTTTTCCCACTATTTTATCATCTCT	1320
Db	1261	CTTTTAAACATAGCAACTCTTTGTAAGTCCCTTTTCCCACTATTTTATCATCTCT	1320
QY	1321	TGAATATGTAATCAGAAATGTTTTTCAAAAACATAAATAAGGTCAAAATAACCGGCT	1380
Db	1321	TGAATATGTAATCAGAAATGTTTTTCAAAAACATAAATAAGGTCAAAATAACCGGCT	1380
QY	1381	ATTTTCAATTTCCATTCACCTTTGTTTCTCAACAATATCAAAACACAGGAATCTCCTTA	1440
Db	1381	ATTTTCAATTTCCATTCACCTTTGTTTCTCAACAATATCAAAACACAGGAATCTCCTTA	1440
QY	1441	TAGTCACTCGCTTTTGTAGTTTCGTCATATG	1470
Db	1441	TAGTCACTCGCTTTTGTAGTTTCGTCATATG	1470
RESULT 2			
CAU64206	2682 bp	DNA	linear
LOCUS	Candida albicans	hyphal wall protein 1 (HWPI)	gene, complete cds
DEFINITION			
ACCESSION	U64206		
VERSION	U64206.1	GI:4028879	
KEYWORDS			
SOURCE	Candida albicans.		
ORGANISM	Candida albicans		
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
AUTHORS	Saccharomycetales; mitosporic Saccharomycetales; Candida.		
TITLE	1 (bases 1 to 2682)		
	Staab, J.F., Ferrer, C.A. and Sundstrom, P.		
	Developmental expression of a tandemly repeated, proline- and		
	glutamine-rich amino acid motif on hyphal surfaces on Candida		
	albicans		
JOURNAL	J. Biol. Chem.	271 (11),	6298-6305 (1996)
MEDLINE	96198091		
PUBMED	8626424		
REFERENCE	2 (bases 1 to 2682)		
AUTHORS	Staab, J.F. and Sundstrom, P.		
TITLE	Genetic organization and sequence analysis of the hypha-specific		
	cell wall protein gene HWPI of Candida albicans		
JOURNAL	Yeast	14 (7),	681-686 (1998)
MEDLINE	98301186		
PUBMED	9639315		
REFERENCE	3 (bases 1 to 2682)		
AUTHORS	Staab, J.F. and Sundstrom, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-JUL-1996)	Medical Microbiology & Immunology, Ohio	
	State University, 333 West 10th Avenue, Columbus, OH 43210, USA		
REFERENCE	4. (bases 1 to 2682)		
AUTHORS	Staab, J.F.		

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

## RESULT 4

CAC35A5  
LOCUS C.albicans cosmid Ca35A5. 42565 bp DNA linear PLN 05-NOV-1998  
ACCESSION AL033396  
VERSION AL033396.1 GI:3850143  
KEYWORDS ARS; Caniki; CDP-alcohol phosphatidyltransferase; chitin synthase; chsi; cytochrome P450; DNA polymerase delta; histidine kinase; isocitrate dehydrogenase pseudogene; LTR; pol3; Rcc1; regulator of chromosome condensation; rehydrin; RPS1-like region.  
SOURCE Candida albicans.  
ORGANISM Candida albicans

## REFERENCE

1 (bases 1 to 42565)  
Tait, E., Simon, M.C., King, S., Brown, A.J., Gow, N.A. and Shaw, D.J.  
A Candida albicans genome project: cosmid contigs, physical mapping, and gene isolation

## JOURNAL

Fungal Genet. Biol. 21 (3), 308-314 (1997)

## MEDLINE

97435544

## REMARK

9290243

## REFERENCE

2 (bases 1 to 42565)

## AUTHORS

Oliver, K. and Harris, D.

## JOURNAL

Unpublished

## REFERENCE

3 (bases 1 to 42565)

## AUTHORS

Barrell, B.G. and Rajandream, M.A.

## TITLE

Direct Submission

## JOURNAL

Submitted (05-NOV-1998)

on the Candida albicans strain 1161 genome. Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. Duncan Shaw, [3] Department of Molecular and Cell Biology, The Institute of Medical Science, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, UK

Notes:  
Funding: sequencing funded by Beowulf Genomics Ltd. CDS are numbered using the following system eg CAC20C1.01c. CA (C. albicans), 20C1 (cosmid name),  
.01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE/PRAM database are also included but some of these may be fortuitous.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid Ca35A5 is likely to map to region R, chromosome 7.

## FEATURES

## source

1..42565  
/organism="Candida albicans"  
/strain="1161"  
/db\_xref="taxon:5476"  
/map="chromosome 7"  
/clone="cosmid Ca35A5"  
complement(1..785)  
/gene="Ca35A5.01c"  
complement(<1..785)  
/gene="Ca35A5.01c"  
note="Ca35A5.01c, partial orf, len: > 261 aa, most similar to EPT1.YEAST ethanolaminephosphotransferase (EC 2.7.8.1) (391 aa), fasta scores opt: 984, E(): 0, (56.3% identity in 263 aa overlap), contains PS00379 CDP-alcohol phosphatidyltransferases signature"  
/codon\_start=1  
/transl\_table=12  
/product="putative alcohol phosphatidyl transferase"  
/protein\_id="CAA21944.1"  
/db\_xref="GI:3850144"  
/translation="MGLFPTNKLONLKLKYSSEHSHIISYILKKWNFVQIFPL SMAPNVTLGLUFFIIGNLMTFFYDPLNFTQPCWCFYAFGLFMVQTFDGCDCGCH ARRTQSGELGFDHSDAINTLTGTFVFAVLKMGYGLLLLSQFASVCNFTSTW EEXYHTLFLSKSGPVEGILMVICIYIITGIFGPDITWIDLFELNLTSLGYGYKVD

## misc\_feature

complement(423..491)  
/gene="Ca35A5.01c"  
/note="PS00379 CDP-alcohol phosphatidyltransferases signature"  
4307..11124

## misc\_feature

/note="RPS1-like region"

## misc\_feature

complement(7606..8085)  
/note="region showing similarity with isocitrate dehydrogenase"

## gene

13175..14713

## CDS

/gene="Ca35A5.02"

13175..14713

/note="Ca35A5.02"

/gene="Ca35A5.02"

/note="Ca35A5.02, unknown, len: 512 aa, possible glycoprotein"

/codon\_start=1

/transl\_table=12

/product="hypothetical protein"

/protein\_id="CAA21945.1"

/db\_xref="GI:3850145"

/translation="MSEVFNQSPSSILKQSNASSSTKVHGAPSSQIANVDFNTLR

KLSNTFSNLNTNTNSALSRFTNRKLSNISVHPSSDDSSPKTLRESTSP

SEKSSVGNRLRIAKLKFQSKNSKPKDLFDTSSTISDSSFRKVIIGNSLNE

TKPRKSSMSPMTTFHSLFRSHSHNSLQDRDNOVATGTTPLGKPDFKSAKTT

LCLSSNSNSIISNPDLAIOYNTNPNISIEDRETNLDHTNSSFDIHKKMLVADSF

IONKLNXYHQTEVGLGIVSELDHNDANKIYNLYLKLPLETSPFSISDSQSKMR

PILSAVEEIANFVKEFCLHQPNIERSPRKTSRSVSLGRDSELDYLLQSLNLF

KMLSLHLDLQTEPEVSLQALILNWKYINAYVRYLLLSIFQPLQIYLNLEFTRSH

NSKIIIRDLLLVSRKRVFTTEQIGSGERETQFLGNAESEDLTGGLLTSLAVL

SSIS"

18031..18456

/gene="Ca35A5.03"

18031..18456

/gene="Ca35A5.03"

/note="Ca35A5.03, unknown, questionable orf, len: 141 aa"

/codon\_start=1

/transl\_table=12

/product="questionable orf"

/protein\_id="CAA21946.1"

/db\_xref="GI:3850146"

/translation="MNQNNKSNTERNSIEEDALHYHYEYSVSLVOHKNKQEIS

MYEIRLKRANSEYRDSIDIKIILVNDQKDSQNEAFSSLMLKSAKQVDKNYOLKI

ENKRLKEIVDAQARLVDFKDEWCFGLKIRKQKV"

19043..19210

/note="BLASTN match (score: 669, strand:(+1), range:

8-175) to AF069450 C. albicans retrotransposon long

terminal repeat zeta"

19204..19277

/note="BLASTN hit (score: 325, strand:(+1), range:

168-24) to AF069450 C. albicans retrotransposon long

terminal repeat zeta"

20359..20769

/note="BLASTN match (score: 1704 strand:(+1), range:

49-459) to CAARESE X65035 C.albicans DNA of an

autonomously replicating sequence (ARS)"

20719..21001

/note="BLASTN match (score: 1298 strand:(+1), range:

405-687) to CAARESE X65035 C.albicans DNA of an

autonomously replicating sequence (ARS)"

20808..23888

/gene="Chs1"

20808..23888

/gene="Chs1"

/note="Chs1"

/gene="Ca35A5.04, Chs1 gene, len: 1026 aa, CHS1\_CANAL

chitin synthase 1 (EC 2.4.1.16), CHS1\_CANAL (776 aa)

(99.9% identity in 776 aa overlap) but open reading frame

here extends 250 aa upstream of CHS1\_CANAL start

at 21561"

/codon\_start=1

/transl\_table=12

/product="chitin synthase"

/protein\_id="CAA21947.1"

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 3, 2003, 19:47:58 ; Search time 15.5 Seconds  
(without alignments)  
94.913 Million cell updates/sec

Title: US-09-725-010-3

Perfect score: 281

Sequence: 1 CTNCTQTTPPLWRNPDPGP.....LHGVRPLSLKTDVIKKNR 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	98.6	853	3	US-09-254-325-2
2	277	98.6	856	3	US-09-079-415-6
3	277	98.6	866	3	US-08-750-458A-2
4	188	66.9	449	2	US-08-927-394-2
5	53	18.9	108	1	US-08-204-740-11
6	53	18.9	108	3	US-09-081-167A-11
7	53	18.9	108	3	US-09-081-395-11
8	53	18.9	108	4	US-09-416-833-11
9	53	18.9	108	5	PCT-US95-02521-11
10	52.5	18.7	2237	1	US-08-455-543A-48
11	52.5	18.7	2237	2	US-08-223-305C-48
12	52.5	18.7	2237	4	US-09-268-163-8
13	52	18.5	128	5	PCT-US95-06266-141
14	49	17.4	314	2	US-08-460-309-19
15	49	17.4	314	2	US-08-125-077-19
16	49	17.4	664	2	US-08-852-153-8
17	49	17.4	786	4	US-09-103-429A-3
18	49	17.4	805	4	US-09-103-429A-4
19	49	17.4	1068	3	US-08-390-874C-11
20	49	17.4	1068	4	US-09-265-772-11
21	49	17.4	1069	2	US-08-162-081B-37
22	49	17.4	1069	2	US-08-780-872-37
23	49	17.4	1069	4	US-09-085-957-37
24	49	17.4	1080	2	US-08-162-081B-36
25	49	17.4	1080	2	US-08-780-872-36
26	49	17.4	1080	4	US-09-085-957-36
27	48.5	17.3	63	1	US-08-269-441A-13

28	48.5	17.3	314	1	US-08-269-441A-2	Sequence 2, Appli
29	48.5	17.3	314	4	US-09-008-465-3	Sequence 3, Appli
30	48.5	17.3	314	4	US-09-276-851-2	Sequence 2, Appli
31	48.5	17.3	314	4	US-09-528-959-3	Sequence 3, Appli
32	48.5	17.3	785	3	US-08-374-077C-3	Sequence 3, Appli
33	48.5	17.3	785	4	US-08-895-590-3	Sequence 3, Appli
34	48.5	17.3	785	4	US-09-539-879A-3	Sequence 3, Appli
35	48.5	17.3	2516	3	US-08-374-077C-2	Sequence 2, Appli
36	48.5	17.3	2516	4	US-08-895-590-2	Sequence 2, Appli
37	48.5	17.3	2516	4	US-09-539-879A-2	Sequence 2, Appli
38	48.5	17.3	2972	4	US-09-579-181-2	Sequence 2, Appli
39	48.5	17.3	3118	4	US-09-579-181-1	Sequence 2, Appli
40	48	17.1	256	2	US-08-484-905-114	Sequence 114, App
41	48	17.1	256	3	US-08-481-985B-114	Sequence 114, App
42	48	17.1	256	4	US-08-370-476-114	Sequence 114, App
43	48	17.1	431	3	US-08-478-507-2	Sequence 2, Appli
44	48	17.1	431	4	US-09-128-275A-2	Sequence 2, Appli
45	48	17.1	431	4	US-09-553-427-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-254-325-2  
; Sequence 2, Application US/09254325  
; Patent No. 6090607  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: ENHANCED EXPRESSION OF  
; TITLE OF INVENTION: PROTEOLYTIC ENZYMES IN KOJI MOLDS  
; NUMBER OF SEQUENCES: 17  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/254,325  
; FILING DATE:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 853 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Binding-site  
; LOCATION: 652-676  
; OTHER INFORMATION: /note= "DNA BINDING SITE"  
; FEATURE:  
; NAME/KEY: Region  
; LOCATION: 1..731  
; OTHER INFORMATION: /note= "TRUNCATED AREA WHICH IS  
; OTHER INFORMATION: STILL ACTIVE BUT NOT REPRESSED BY  
; OTHER INFORMATION: L-GLUTAM..."  
US-09-254-325-2

Query Match 98.6%; Score 277; DB 3; Length 853;  
Best Local Similarity 98.0%; Pred. No. 1.3e-30;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCTQTTPPLWRNPDPGPICNAGGLFLKHGVVRPLSLKTDVIKKNR 50  
|||||  
Db 652 CTNCTQTTPPLWRNPDPGPICNAGGLFLKHGVVRPLSLKTDVIKKNR 701

RESULT 2  
US-09-079-415-6  
; Sequence 6, Application US/09079415  
; Patent No. 6013452  
; GENERAL INFORMATION:

; APPLICANT: Christensen, Tove  
; APPLICANT: Lehbeck, Jan  
; TITLE OF INVENTION: A Fungus Wherein The area, pepC and/or  
; TITLE OF INVENTION: pepE Genes Have Been Inactivated  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 60134520 No. 6013452disk of No. 6013452th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/079,415  
; FILING DATE: 14-MAY-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rozek, Carol  
; REGISTRATION NUMBER: 36,993  
; REFERENCE/DOCKET NUMBER: 4657-204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 866 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-079-415-6

Query Match 98.6%; Score 277; DB 3; Length 866;  
Best Local Similarity 98.0%; Pred. No. 1.4e-30;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50  
|||||  
DB 664 CTNCFQTTPPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVIKKNR 713

RESULT 3  
US-08-750-458A-2  
; Sequence 2, Application US/08750458A  
; Patent No. 6025185  
; GENERAL INFORMATION:  
; APPLICANT: Christensen, Tove  
; APPLICANT: Hynes, Michael J.  
; TITLE OF INVENTION: A Fungus Wherein The Area Gene Has Been Modified And An Area  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 60251850 No. 6025185disk of No. 6025185th America, Inc.  
; STREET: 405 Lexington Avenue, Suite 6400  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10174-6401  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/750,458A  
; FILING DATE: 3-December-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rozek, Carol E.  
; REGISTRATION NUMBER: 36,993

; REFERENCE/DOCKET NUMBER: 4129-204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 867 0123  
; TELEFAX: 212 867 0298  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 866 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-750-458A-2

Query Match 98.6%; Score 277; DB 3; Length 866;  
Best Local Similarity 98.0%; Pred. No. 1.4e-30;  
Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50  
|||||  
DB 664 CTNCFQTTPPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVIKKNR 713

RESULT 4  
US-08-927-394-2  
; Sequence 2, Application US/08927394  
; Patent No. 5990092  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: GATA-6 TRANSCRIPTION FACTOR: COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/927,394  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/025,574  
; FILING DATE: 06-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Plumer, Elizabeth R.  
; REGISTRATION NUMBER: 36,637  
; REFERENCE/DOCKET NUMBER: S1237/7005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 449 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-927-394-2

Query Match 66.9%; Score 188; DB 2; Length 449;  
Best Local Similarity 62.0%; Pred. No. 2.1e-18;  
Matches 31; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDPGLCNACGLFLKLGHVVRPLSLKTDVIKKNR 50  
|||||  
DB 298 CANCHTTTTLWRNAGEPVCNACGLYMKLHGVPRLAMKKEGIQTRR 347

RESULT 5



US-08-204-740-11  
; Sequence 11, Application US/08204740  
; Patent No. 5753432

; GENERAL INFORMATION:  
; APPLICANT: Gudkov, Andrei

; APPLICANT: Kazarov, Alexander  
; APPLICANT: Mazo, Ilya

; APPLICANT: Roninson, Igor B  
; TITLE OF INVENTION: Methods for Identifying Genetic

; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant

; TITLE OF INVENTION: Growth in Cancer Cells  
; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Allgretti & Witcoff, Ltd.

; STREET: 10 S. Wacker Drive, Suite 3000  
; CITY: Chicago

; STATE: Illinois  
; COUNTRY: USA

; ZIP: 60606  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
; FILING DATE: 04-MAR-1994

; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: No. 5753432nan, Kevin E  
; REGISTRATION NUMBER: 35,303

; REFERENCE/DOCKET NUMBER: 93,354-C  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234

; TELEX: 910-221-5317  
; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 108 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear

; MOLECULE TYPE: peptide  
US-08-204-740-11

Query Match 18.9%; Score 53; DB 1; Length 108;

Best Local Similarity 43.5%; Pred. No. 3.8;

Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 28 FLKLHGVRPLSLKTDVKKRNR 50

Db 14 FLEANEVPRPVLTNTLTTRR 36

RESULT 6  
US-09-081-167A-11

; Sequence 11, Application US/09081167A  
; Patent No. 6083745

; GENERAL INFORMATION:  
; APPLICANT: Gudkov, Andrei

; APPLICANT: Kazarov, Alexander  
; APPLICANT: Mazo, Ilya

; APPLICANT: Roninson, Igor B  
; TITLE OF INVENTION: Methods for Identifying Genetic

; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant

; TITLE OF INVENTION: Growth in Cancer Cells  
; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

; STREET: 300 S. Wacker Drive, 32nd Floor  
; CITY: Chicago

; STATE: Illinois  
; COUNTRY: USA

; ZIP: 60606  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
; FILING DATE: 18-MAY-1998

; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: No. 6083746nan, Kevin E  
; REGISTRATION NUMBER: 35,303

; REFERENCE/DOCKET NUMBER: 93,354-KK  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002

; TELEX:  
; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 108 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

; FILING DATE: 18-MAY-1998  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6083745nan, Kevin E

; REGISTRATION NUMBER: 35,303  
; REFERENCE/DOCKET NUMBER: 93,354-KK

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001

; TELEFAX: 312-913-0002  
; TELEX:

; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 108 amino acids  
; TYPE: amino acid

; TOPOLOGY: linear  
; MOLECULE TYPE: peptide

US-09-081-167A-11

Query Match 18.9%; Score 53; DB 3; Length 108;

Best Local Similarity 43.5%; Pred. No. 3.8;

Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 28 FLKLHGVRPLSLKTDVKKRNR 50

Db 14 FLEANEVPRPVLTNTLTTRR 36

RESULT 7  
US-09-081-395-11

; Sequence 11, Application US/09081395  
; Patent No. 6083746

; GENERAL INFORMATION:  
; APPLICANT: Gudkov, Andrei

; APPLICANT: Kazarov, Alexander  
; APPLICANT: Mazo, Ilya

; APPLICANT: Roninson, Igor B  
; TITLE OF INVENTION: Methods for Identifying Genetic

; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant

; TITLE OF INVENTION: Growth in Cancer Cells  
; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

; STREET: 300 S. Wacker Drive, 32nd Floor  
; CITY: Chicago

; STATE: Illinois  
; COUNTRY: USA

; ZIP: 60606  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
; FILING DATE: 18-MAY-1998

; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

; NAME: No. 6083746nan, Kevin E  
; REGISTRATION NUMBER: 35,303

; REFERENCE/DOCKET NUMBER: 93,354-KK  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002

; TELEX:  
; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 108 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear

```

; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-081-395-11

```

Query Match 18.9%; Score 53; DB 3; Length 108;  
Best Local Similarity 43.5%; Pred. No. 3.8;  
Matches 10; Conservative 7; Mismatches 6; Indels

```
Qy 28 FLKLGVRPLSLKTDVIKKRNR 50
    ||: : | ||: |: |: | | |
Db 14 FLEANEVPRPVTLRNTLKTTRRR 36
```

## RESULT 8

US-09-416-833-11  
; Sequence 11, Application US/09416833  
; Patent No. 6197521

APPLICANT: Gudkov, Andrei  
APPLICANT: Kazarov, Alexander  
APPLICANT: Mazo, Ilya  
APPLICANT: Roninson, Igor B

AFFRICANI, ROLANDSON, 1981 B  
 TITLE OF INVENTION: Methods for Identifying Genetic  
 TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant  
 TITLE OF INVENTION: Growth in Cancer Cells  
 NUMBER OF SEQUENCES: 13

NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allgretti & Witcoff, Ltd.  
STREET: 10 S. Wacker Drive, Suite 3000

```

; ZLF: 00000
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-
; SOFTWARE: PatentIn Release t
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/4

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/204,740
; FILING DATE:

```

ATTORNEY/AGENT INFORMATION:  
NAME: NO. 6197521nan, Kevin E  
REGISTRATION NUMBER: 35,303  
REFERENCE/DOCKET NUMBER: 93,354-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-715-1000  
TELEFAX: 312-715-1234  
TELEX: 910-221-5317

```

; ELEA: 910-421-3517
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-416-833-11

```

Query Match 18.9%; Score 53; DB 4; Length 108;  
Best Local Similarity 43.5%; Pred. No. 3.8;  
Matches 10: Conservative 7: Mismatches 6: Indels

QY	28	FLKLHGVRPLSLKTDVIKRRNR	50
		: :     :  :	
Db	14	FLEANEVPRPVTLRNTLTKRRR	36

## RESULT 9

```

PCT-US95-02521-11
; GENERAL INFO: Application PC/TUS9502521
; SERIAL INFORMATION:
;
; APPLICANT:
;
; TITLE OF INVENTION: Methods for Identifying
; TITLE OF INVENTION: Suppressor Elements
; TITLE OF INVENTION: Growth in Cancer
; NUMBER OF SEQUENCES: 13
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, V01.00
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02521
; FILING DATE:

```

```

; PDB ID: 1PCT
; FILE: 1PCT.DAT
; INFORMATION FOR SEQ ID NO: 11:
;-----
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 108 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US95-02521-11

```

```
Query Match      18.9%; Score 53; DB 5; Length 108;
Best Local Similarity 43.5%; Pred. No. 3.8;
Matches 10: Conservative 7; Mismatches 6; Indels 0; Gaps 0;
```

Qy 28 FLKLHGVRPLSLKTDVIKRRN 50  
||: : | ||: |: : | | |  
Db 14 FLEANEVPRPVTLRTNTLKTRRR 36

## RESULT 10

US-08-455-543A-48  
; Sequence 48, Application US/08455543A  
; Patent No. 5792846

APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: Feldman, Daniel  
APPLICANT: McCue, Ann

APPLICANT: Brenner, Robert  
 TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND METHODS  
 NUMBER OF SEQUENCES: 57  
 CORRESPONDENCE ADDRESS:

; CORRESPONDENCE ADDRESS.  
 ; ADDRESSEE: Brown, Martin, Haller & McClain  
 ; STREET: 1660 Union Street  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: USA  
 ;

```

? COUNTING LOG
? ZIP: 91010-2926
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSEQ Version 1.5
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/455,
? FILING DATE: May 31, 1995
?
```

? PRIOR APPLICATION DATA: 08/223,305  
 ? APPLICATION NUMBER:  
 ? FILING DATE: April 4, 1994  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER: 07/868,354  
 ? FILING DATE: April 10, 1992  
 ? PRIOR APPLICATION DATA: US 07/745,206  
 ? APPLICATION NUMBER:  
 ? FILING DATE: 15-AUG-1991  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER: US 07/620,250

;; FILING DATE: 30-NOV-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/482,384  
;; FILING DATE: 20-FEB-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/603,751  
;; FILING DATE: 04-APR-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: WO PCT/US89/01408  
;; FILING DATE: 04-APR-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/176,899  
;; FILING DATE: 04-APR-1988  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie L.  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 6362-52517  
;; TELEPHONE: (619)238-0999  
;; TELEFAX: (619)238-0062  
;; INFORMATION FOR SEQ ID NO: 48:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2237 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; FRAGMENT TYPE: internal  
;; US-08-455-543A-48

Query Match 18.7%; Score 52.5; DB 1; Length 2237;  
Best Local Similarity 37.3%; Pred. No. 1.7e+02;  
Matches 19; Conservative 3; Mismatches 22; Indels 7; Gaps 2;

QY 5 FTQTPTLWRNPDGQPLCNACGLFLKLH----GVVRPLS---LKTDTVKKR 48  
Db 2170 FPNTPCCRETSPASWPPLALELALTLTGWSVWTVRPLSTPCLTRLSRR 2220

RESULT 11  
US-08-223-305C-48  
; Sequence 48, Application US/08223305C  
; Patent No. 5851824  
; GENERAL INFORMATION:  
; APPLICANT: Harpold, Michael  
; APPLICANT: Ellis, Steven  
; APPLICANT: Williams, Mark  
; APPLICANT: Feldman, Daniel  
; APPLICANT: McCue, Ann  
; APPLICANT: Brenner, Robert  
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/223,305C  
; FILING DATE: April 4, 1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/868,354  
; FILING DATE: April 10, 1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/745,206

;; FILING DATE: 15-AUG-1991  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/620,250  
;; FILING DATE: 30-NOV-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/482,384  
;; FILING DATE: 20-FEB-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/603,751  
;; FILING DATE: 04-APR-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: WO PCT/US89/01408  
;; FILING DATE: 04-APR-1989  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/176,899  
;; FILING DATE: 04-APR-1988  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Seidman, Stephanie L.  
;; REGISTRATION NUMBER: 33,779  
;; REFERENCE/DOCKET NUMBER: 52516 (P519739)  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619)238-0999  
;; TELEFAX: (619)238-0062  
;; INFORMATION FOR SEQ ID NO: 48:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2237 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; FRAGMENT TYPE: internal  
;; US-08-223-305C-48

Query Match 18.7%; Score 52.5; DB 2; Length 2237;  
Best Local Similarity 37.3%; Pred. No. 1.7e+02;  
Matches 19; Conservative 3; Mismatches 22; Indels 7; Gaps 2;

QY 5 FTQTPTLWRNPDGQPLCNACGLFLKLH----GVVRPLS---LKTDTVKKR 48  
Db 2170 FPNTPCCRETSPASWPPLALELALTLTGWSVWTVRPLSTPCLTRLSRR 2220

RESULT 12  
US-09-268-163-8  
; Sequence 8, Application US/09268163B  
; Patent No. 6353091  
; GENERAL INFORMATION:  
; APPLICANT: Lipscombe, Diane  
; APPLICANT: Schorge, Stephanie  
; TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISOFORM AND USES THEREOF  
; FILE REFERENCE: B1055/7000  
; CURRENT APPLICATION NUMBER: US/09/268,163B  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 60/077,901  
; EARLIER FILING DATE: 1998-03-13  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 8  
; LENGTH: 2237  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-268-163-8

Query Match 18.7%; Score 52.5; DB 4; Length 2237;  
Best Local Similarity 37.3%; Pred. No. 1.7e+02;  
Matches 19; Conservative 3; Mismatches 22; Indels 7; Gaps 2;

QY 5 FTQTPTLWRNPDGQPLCNACGLFLKLH----GVVRPLS---LKTDTVKKR 48  
Db 2170 FPNTPCCRETSPASWPPLALELALTLTGWSVWTVRPLSTPCLTRLSRR 2220

RESULT 13

```

; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,309
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,077
; FILING DATE: 22-SEP-1993
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-460-309-19

Query Match 17.4%; Score 49; DB 2; Length 314;
Best Local Similarity 27.3%; Pred. No. 50;
Matches 15; Conservative 4; Mismatches 12; Indels 24;

QY 1 CTCN---FTQTTPLWRNPQGPL-CNAC-----GLFLKLHG 33
Db 35 CNECCGCFEKK--WRQNTNARFNCPCNCHGHSNECKYDEVNRRKGLSLDING 87

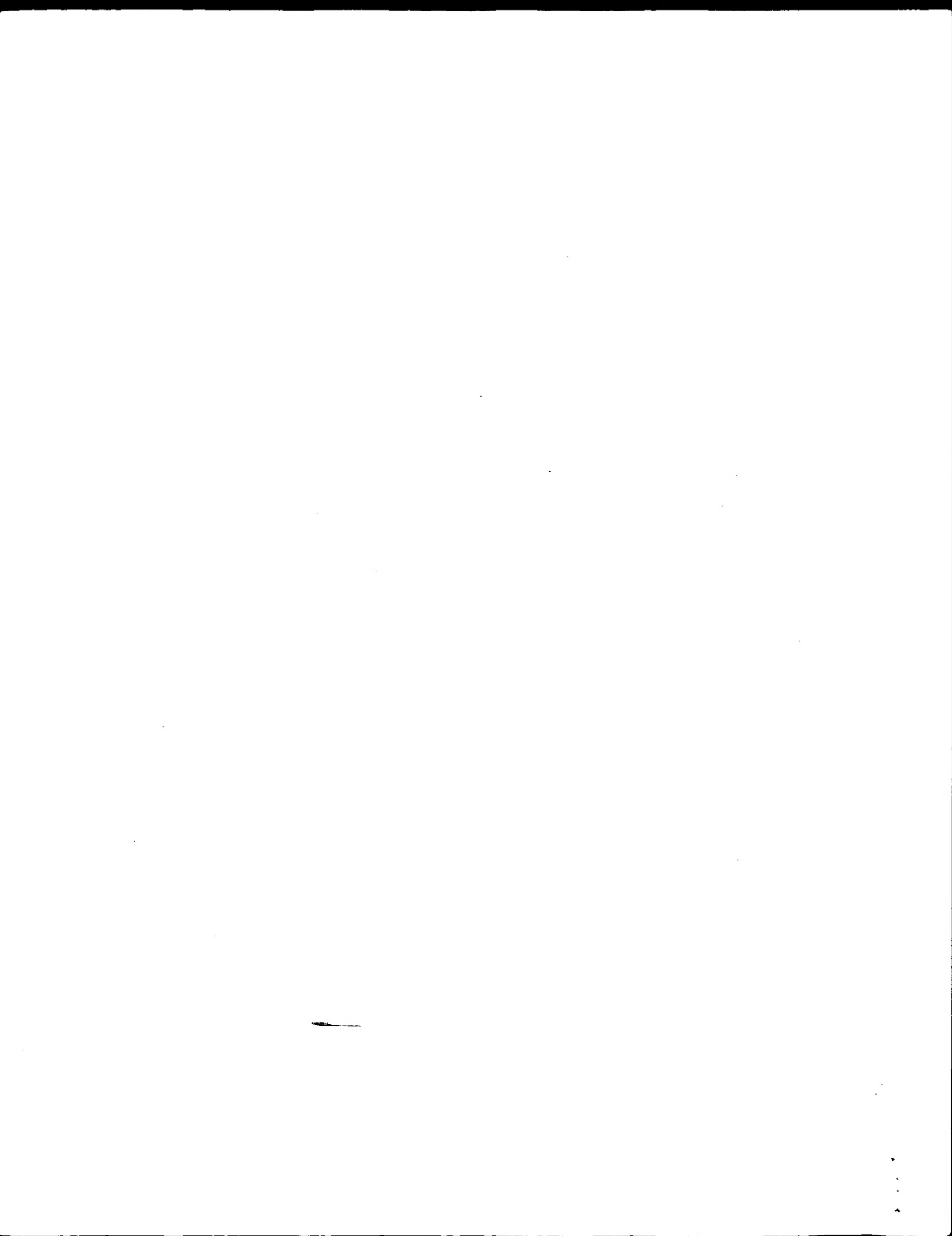
RESULT 15
US-08-125-077-19
; Sequence 19, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA

```

ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/125,077  
FILING DATE: 22-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US 94/10730  
FILING DATE: 21-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/472,319  
FILING DATE: 30-JAN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/919,951  
FILING DATE: 27-JUL-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LA 9721  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 314 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-125-077-19

Query Match 17.4%; Score 49; DB 2; Length 314;  
Best Local Similarity 27.3%; Pred. No. 50;  
Matches 15; Conservative 4; Mismatches 12; Indels 24; Gaps 4;  
QY 1 CTNC---FTQTPLWRRNPDGQPL-CNAC-----GLFLKLHG 33  
Db 35 CNECCPGFEQKK--WRQNTNARPNCEPCNCHGHSNECKYDEEVNKGSLDING 87

Search completed: January 3, 2003, 19:52:38  
Job time : 16.5 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:46:47 ; Search time 25 Seconds  
(without alignments)  
192.269 Million cell updates/sec

Title: US-09-725-010-2  
Perfect score: 279  
Sequence: 1 CTNGCTKTPLEWRNPOGOP.....LHGVRPLSLKTDVIKKRQR 50

Scoring table: BLOSUM62  
Gapop 10.0 . Gapext 0.5

```
Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224
```

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

## SUMMARTES

Result No.	Query %			Length	DB	ID	Description
	Score	Match					
1	258	92.5	862	2	S51493	major nitrogen reg	
2	258	92.5	876	1	A57988	regulatory protein	
3	256	91.8	1036	1	A34755	nitrogen regulator	
4	242	86.7	855	2	T41336	probable nitrogen	
5	238	85.3	510	2	S56233	probable membrane	
6	202	72.4	730	2	S50543	GN3 protein - yea	
7	201	72.0	551	2	S53377	GUF3 protein - yea	
8	195	69.9	388	2	I51420	transcription fact	
9	195	69.9	390	2	I51419	transcription fact	
10	194	69.5	532	2	JC6170	GATA-transcription	
11	191	68.5	779	1	S40382	box A-binding fact	
12	190	68.1	269	2	S22781	transcription fact	
13	189	67.7	391	2	I50702	transcription fact	
14	188	67.4	387	2	I50703	transcription fact	
15	188	67.4	391	2	A49645	transcription fact	
16	185	66.3	380	2	I50701	transcription fact	
17	185	66.3	439	1	A48099	transcription fact	
18	185	66.3	439	2	I57561	transcription fact	
19	185	66.3	440	2	I61183	transcription fact	
20	184	65.9	564	2	T38291	GATA-type transcri	
21	183	65.6	950	2	S27473	URB51 protein - sm	
22	178	63.8	241	2	S53812	BmGATA beta isofor	
23	178	63.8	327	2	S53811	BmGATA beta isofor	
24	178	63.8	413	2	S04655	transcription fact	
25	178	63.8	509	2	A53741	transcription fact	
26	178	63.8	564	2	T43298	transcription fact	
27	177	63.4	304	2	A32993	transcription fact	
28	177	63.4	413	1	A34888	transcription fact	
29	177	63.4	413	2	S48756	transcription fact	

## ALIGNMENTS

## RESULT 1

major nitrogen regulation protein - *Penicillium chrysogenum* (strain Q176)  
 C:Species: *Penicillium chrysogenum*  
 A:Variatey: strain Q176  
 C:Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 08-Dec-2000  
 C:Accession: S51493  
 R:Haas, H.; Bauer, B.; Redl, B.; Stoeffler, G.; Marzluf, G.A.  
 Curr. Genet. 27, 150-158, 1995  
 A:title: Molecular cloning and analysis of nre, the major nitrogen regulatory gene of  
 A:Reference number: S51493; MUID:95308537; PMID:7788718  
 A:Accession: S51493  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-862 <HAA>  
 A:Cross-references: EMBL:U02612  
 C:Genetics:  
 A:Introns: 123/2  
 C:Superfamily: nitrogen regulatory protein nit-2; GATA-type zinc finger homology  
 C:Keywords: zinc finger  
 F:659-712/Domain: GATA-type zinc finger homology <GZF>

[illegible]

## RESULT 2

A57988  
 regulatory protein *areA* - *Emericella nidulans*  
 C:Species: *Emericella nidulans*, *Aspergillus nidulans*  
 C:Date: 26-Jul-1996 #sequence.revision 26-Jul-1996 #text\_change 16-Jun-2000  
 C:Accession: A57988; S10017; S70168; S72883  
 R:Kudla, B.; Caddick, M.X.; Langdon, T.; Martinez-Rossi, N.M.; Bennett, C.F.; Sibley, E.M.B.O. J. 9, 1355-1364, 1990  
 A:Title: The regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus*  
 A:Reference number: S10017; MUID:90228331; PMID:1570293  
 A:Accession: A57988  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-876 <KU2>  
 A:Cross-references: EMBL:X52491; NID:g1019911; PIDN:CAA36731.1; PID:g1154625  
 A:Note: this sequence represents reinterpretation to include two exons  
 A:Accession: S10017  
 A:Molecule type: DNA  
 A:Residues: 158-876 <KU2>  
 A:Cross-references: EMBL:X52491  
 A:Note: this sequence represents the authors' original translation





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2003, 19:42:57 ; Search time 48 Seconds  
(without alignments)  
214.632 Million cell updates/sec

Title: US-09-725-010-3

Perfect score: 281

Sequence: 1 CTNCFQTTPPLWRNPDGQP.....LHGVRPLSLKTDVIKRN 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	98.6	853	3 Q9C414	Q9C414 aspergillus
2	277	98.6	866	3 Q9Y7E8	Q9Y7E8 aspergillus
3	277	98.6	918	3 Q9HEW7	Q9HEW7 cladosporiu
4	277	98.6	944	3 Q60043	Q60043 metarhizium
5	247	87.9	1075	3 Q9P952	Q9P952 schizoshyll
6	240	85.4	327	3 P87016	P87016 zygosacchar
7	204	72.6	298	3 Q59842	Q59842 penicillium
8	198	70.5	312	3 Q9HEV2	Q9HEV2 emericella
9	198	70.5	320	3 Q9HEV3	Q9HEV3 emericella
10	198	70.5	436	3 Q9HEV4	Q9HEV4 emericella
11	195	69.4	383	13 Q9W6U0	Q9W6U0 brachydanio
12	194	69.0	119	5 Q77157	Q77157 strongyloce
13	194	69.0	537	5 Q95VY5	Q95VY5 anopheles g
14	193	68.7	486	5 Q8T3J0	Q8T3J0 drosophila
15	190	67.6	324	5 Q9NJ18	Q9NJ18 ceratitidis c
16	190	67.6	511	5 Q8WQ44	Q8WQ44 calliphora

17	190	67.6	949	5 Q94884	Q94884 drosophila
18	190	67.6	950	5 Q9VF01	Q9VF01 drosophila
19	188	66.9	418	13 Q90410	Q90410 brachydanio
20	188	66.9	446	6 Q95JA5	Q95JA5 sus scrofa
21	185	65.8	338	13 Q9PTJ2	Q9PTJ2 brachydanio
22	184	65.5	383	13 Q9PTJ1	Q9PTJ1 brachydanio
23	183	65.1	266	6 Q95KV6	Q95KV6 bos taurus
24	183	65.1	431	5 Q77156	Q77156 strongyloce
25	183	65.1	439	4 Q16365	Q16365 homo sapien
26	183	65.1	442	11 Q8VI87	Q8VI87 mus musculu
27	182	64.8	116	6 Q9MYT2	Q9MYT2 oryctolagus
28	179	63.7	118	11 Q9QW52	Q9QW52 mus sp. ret
29	179	63.7	480	4 Q96BH0	Q96BH0 homo sapien
30	179	63.7	480	11 Q9DC59	Q9DC59 mus musculu
31	179	63.7	480	11 Q9DBY9	Q9DBY9 mus musculu
32	179	63.7	480	11 Q924I4	Q924I4 rattus norv
33	179	63.7	480	11 Q8VHY4	Q8VHY4 rattus norv
34	178	63.3	194	13 Q90411	Q90411 brachydanio
35	178	63.3	444	11 Q99NH5	Q99NH5 rattus norv
36	178	63.3	444	13 Q8QFW5	Q8QFW5 raja eglant
37	178	63.3	549	3 Q9Y754	Q9Y754 emericella
38	176	62.6	704	5 Q9BPQ3	Q9BPQ3 drosophila
39	176	62.6	734	5 Q9VF00	Q9VF00 drosophila
40	175	62.3	388	3 Q9C154	Q9C154 botrytis ci
41	173	61.6	868	5 Q9NFF4	Q9NFF4 aedes aegyp
42	171	60.9	587	3 Q74262	Q74262 neurospora
43	165	58.7	232	5 Q8SUM6	Q8SUM6 encephalito
44	160	56.9	352	5 Q9BPQ1	Q9BPQ1 drosophila
45	158	56.2	466	4 Q96BH8	Q96BH8 homo sapien

## ALIGNMENTS

RESULT 1

Q9C414 ID Q9C414 PRELIMINARY; PRT; 853 AA.

AC Q9C414; DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Area.

GN AREA.

OS Aspergillus oryzae.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OX NCBI\_TaxID=5062;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TK3;

RA van den Broek P.;

RT "The A. oryzae area gene as a tool to overproduce a large number of

proteases.";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF320305; AAK08066.1; .

DR HSSP; P17429; 4GAT.

DR InterPro; IPR000679; Znf\_GATA.

DR Pfam; PF00320; GATA; 1.

DR PRINTS; PR00619; GATAZNFINGER.

DR SMART; SM00401; Znf\_GATA; 1.

DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.

DR PROSITE; PS01114; GATA\_ZN\_FINGER\_2; 1.

SQ SEQUENCE 853 AA; 91497 MW; A5EC8ICE256A6B15 CRC64;

Query Match

Best Local Similarity 98.6%; Score 277; DB 3; Length 853;

Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFQTTPPLWRNPDGQPPLCNACGLFLKLHGVRPLSLKTDVIKRN 50

|||||

DB 652 CTNCFQTTPPLWRNPDGQPPLCNACGLFLKLHGVRPLSLKTDVIKRN 701

## RESULT 2

Q9Y7E8 PRELIMINARY; PRT; 866 AA.  
 AC Q9Y7E8;  
 DT 01-NOV-1999 (TReMBLrel. 12, Created)  
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE Major nitrogen regulatory protein.  
 GN AREA.  
 OS Aspergillus parasiticus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 56775;  
 RX MEDLINE=20225467; PubMed=10760588;  
 RA Chang P.-K., Yu J., Bhatnagar D., Cleveland T.E.;  
 RT "Characterization of Aspergillus parasiticus major nitrogen regulatory  
 gene, area";  
 RL Biochim. Biophys. Acta 1491:263-266(2000).  
 DR EMBL; AF148539; AAD37409.1; -;  
 DR HSSP; P17429; 4GAT.  
 DR InterPro; IPR000679; Znf\_GATA.  
 DR Pfam; PF00320; GATA; 1.  
 DR PRINTS; PR00619; GATAZNFINGER.  
 DR SMART; SM00401; Znf\_GATA; 1.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
 DR PROSITE; PS50114; GATA\_ZN\_FINGER\_2; 1.  
 SQ SEQUENCE 866 AA; 92877 MW; 16FF4D5E823988 CRC64;

Query Match 98.6%; Score 277; DB 3; Length 866;  
 Best Local Similarity 98.0%; Pred. No. 8.7e-30;  
 Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50  
 |||||  
 Db 664 CTNCFTQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 713

## RESULT 3

Q9HEW7 PRELIMINARY; PRT; 918 AA.  
 AC Q9HEW7;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Nitrogen response factor NRF1.  
 OS Cladosporium fulvum (Fulvia fulva).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;  
 OC Dothideomycetes et Chaetothyriomycetes incertae sedis;  
 OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.  
 OX NCBI\_TaxID=5499;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Perez-Garcia A., Snoeijers S.S., Joosten M.H.A.J., Goosen T.,  
 RA De Wit P.J.G.M.;  
 RT "Expression of the avirulence gene Avr9 of the fungal tomato pathogen  
 Cladosporium fulvum is regulated by the global nitrogen response  
 factor NRF1";  
 RL Mol. Plant Microbe Interact. 0:0-0(2001).  
 DR EMBL; AF312694; AAG48616.1; -;  
 DR HSSP; P17429; 4GAT.  
 DR InterPro; IPR002965; P-rich\_extensn.  
 DR InterPro; IPR000679; Znf\_GATA.  
 DR Pfam; PF00320; GATA; 1.  
 DR PRINTS; PR00619; GATAZNFINGER.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR SMART; SM00401; Znf\_GATA; 1.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
 DR PROSITE; PS50114; GATA\_ZN\_FINGER\_2; 1.  
 SQ SEQUENCE 918 AA; 99077 MW; E98198D999BCA899 CRC64;

Query Match 98.6%; Score 277; DB 3; Length 918;  
 Best Local Similarity 98.0%; Pred. No. 9.3e-30;  
 Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50  
 |||||  
 Db 673 CTNCFTQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 722

## RESULT 4

O60043 PRELIMINARY; PRT; 944 AA.  
 AC O60043;  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE Nitrogen response regulator.  
 GN NRRL.  
 OS Metarhizium anisopliae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.  
 OX NCBI\_TaxID=5530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MEL;  
 RX MEDLINE=99070047; PubMed=9852945;  
 RA Screen S.E., Bailey A.M., Charnley K., Cooper R., Clarkson J.M.;  
 RT "Isolation of a nitrogen response regulator gene (nrr1) from  
 Metarhizium anisopliae";  
 RL Gene 221:17-24(1998).  
 DR EMBL; AJ006468; CAA07052.1; -;  
 DR HSSP; P17429; 4GAT.  
 DR InterPro; IPR000679; Znf\_GATA.  
 DR Pfam; PF00320; GATA; 1.  
 DR PRINTS; PR00619; GATAZNFINGER.  
 DR SMART; SM00401; Znf\_GATA; 1.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
 DR PROSITE; PS50114; GATA\_ZN\_FINGER\_2; 1.  
 SQ SEQUENCE 944 AA; 99687 MW; ALA723E658C23EAO CRC64;

Query Match 98.6%; Score 277; DB 3; Length 944;  
 Best Local Similarity 98.0%; Pred. No. 9.6e-30;  
 Matches 49; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNCFTQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 50  
 |||||  
 Db 663 CTNCFTQTTPLRNPDGQPLCNACGLFLKLGWVRPLSLKTDVIKKNR 712

## RESULT 5

Q9P952 PRELIMINARY; PRT; 1075 AA.  
 AC Q9P952;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Transcription factor ScGATA-6.  
 GN SCGT-6.  
 OS Schizophyllum commune (Bracket fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Schizophyllaceae; Schizophyllum.  
 OX NCBI\_TaxID=5334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamagishi K.;  
 RT "putative transcription factor ScGATA-6";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB043467; BAA96108.1; -;  
 DR HSSP; P17429; 4GAT.  
 DR InterPro; IPR000679; Znf\_GATA.  
 DR Pfam; PF00320; GATA; 1.  
 DR PRINTS; PR00619; GATAZNFINGER.  
 DR SMART; SM00401; Znf\_GATA; 1.

DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
DR PROSITE; PS00114; GATA\_ZN\_FINGER\_2; 1.  
SQ SEQUENCE 1075 AA; 113322 MW; D29343211C02FFEA CRC64;  
Query Match 87.9%; Score 247; DB 3; Length 1075;  
Best Local Similarity 88.0%; Pred. No. 1.6e-25;  
Matches 44; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVKKRNR 50  
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 957 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVKKRNR 1006  
RESULT 6  
P87016 ID P87016 PRELIMINARY; PRT; 327 AA.  
AC P87016;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE DNA-binding protein.  
GN SAT1.  
OS Zygosaccharomyces rouxii (Candida mogii).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.  
OX NCBI\_TaxID=4956;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ushio K., Otsuka H., Yoshikawa S., Taguchi G., Shimosaka M.,  
RA Mitsu N., Okazaki M.;  
RT "Cloning of the SAT1 gene concerned with salt tolerance of the yeast  
RT Zygosaccharomyces rouxii.";  
RL J. Ferment. Bioeng. 82:16-21(1996).  
DR EMBL; D83211; BAA18948.1; -;  
DR HSP; P17429; 4GAT.  
DR InterPro; IPR000679; Znf\_GATA.  
DR Pfam; PF00320; GATA; 1.  
DR PRINTS; PR00619; GATAZNFINGER.  
DR SMART; SM00401; Znf\_GATA; 1.  
DR PROSITE; PS00114; GATA\_ZN\_FINGER\_2; 1.  
KW DNA-binding  
SQ SEQUENCE 327 AA; 37304 MW; 9D7AF1002807283F CRC64;  
Query Match 85.4%; Score 240; DB 3; Length 327;  
Best Local Similarity 86.0%; Pred. No. 4.3e-25;  
Matches 43; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVKKRNR 50  
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 246 CTHCHTRTPPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVKKRQR 295  
RESULT 7  
O59842 ID O59842 PRELIMINARY; PRT; 298 AA.  
AC O59842;  
DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GATA transcription factor.  
GN NREB.  
OS Penicillium chrysogenum.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
OX NCBI\_TaxID=5076;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Haas H., Angermayr K., Zadra I.;  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U96385; AAC09045.1; -;  
DR HSP; P17429; 4GAT.  
DR InterPro; IPR000679; Znf\_GATA.  
DR Pfam; PF00320; GATA; 1.

DR PRINTS; PR00619; GATAZNFINGER.  
DR SMART; SM00401; Znf\_GATA; 1.  
DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
DR PROSITE; PS00114; GATA\_ZN\_FINGER\_2; 1.  
SQ SEQUENCE 298 AA; 32670 MW; C220D36F4993C590 CRC64;  
Query Match 72.6%; Score 204; DB 3; Length 298;  
Best Local Similarity 76.0%; Pred. No. 3.9e-20;  
Matches 38; Conservative 2; Mismatches 10; Indels 0; Gaps 0;  
QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVKKRNR 50  
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 17 CQCGTSTPLWRNDELGSVLCNACGLFLKLGHRPRPISLKTVDVKSRR 66  
RESULT 8  
Q9HEV2 ID Q9HEV2 PRELIMINARY; PRT; 312 AA.  
AC Q9HEV2;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE GATA factor AREB alpha.  
GN AREB.  
OS Emericella nidulans (Aspergillus nidulans).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; Emericella.  
OX NCBI\_TaxID=5072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89218944; PubMed=2651886;  
RA Arst H.N. Jr., Tollervey D., Caddick M.X.;  
RT "A translocation associated, loss-of-function mutation in the nitrogen  
RT metabolite repression regulatory gene of Aspergillus nidulans can  
RT revert intracristically.";  
RL Mol. Gen. Genet. 215:364-367(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21206224; PubMed=11309119;  
RA Conlon H., Zadra I., Haas H., Arst H.N. Jr., Jones M.G., Caddick M.X.;  
RT "The Aspergillus nidulans GATA transcription factor gene areb encodes  
RT at least three proteins and features three classes of mutation.";  
RL Mol. Microbiol. 40:361-375(2001).  
DR EMBL; AF320976; AAG49353.1; -;  
DR HSP; P17429; 4GAT.  
DR InterPro; IPR000679; Znf\_GATA.  
DR Pfam; PF00320; GATA; 1.  
DR PRINTS; PR00619; GATAZNFINGER.  
DR SMART; SM00401; Znf\_GATA; 1.  
DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
DR PROSITE; PS00114; GATA\_ZN\_FINGER\_2; 1.  
SQ SEQUENCE 312 AA; 34279 MW; 35A73B7526FE5F88 CRC64;  
Query Match 70.5%; Score 198; DB 3; Length 312;  
Best Local Similarity 74.0%; Pred. No. 2.8e-19;  
Matches 37; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
QY 1 CTNCFQTTPPLWRNPDGQPLCNACGLFLKLGWVRPLSLKTDVKKRNR 50  
||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 17 CQCGTSTPLWRNDELGSVLCNACGLFLKLGHRPRPISLKTVDVKSRR 66  
RESULT 9  
Q9HEV3 ID Q9HEV3 PRELIMINARY; PRT; 320 AA.  
AC Q9HEV3;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE GATA factor AREB beta.  
GN AREB.  
OS Emericella nidulans (Aspergillus nidulans).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;



Q8T3J0	PRELIMINARY;	PRT;	486 AA.
Q8T3J0	ID	Q8T3J0	Created
AC	Q8T3J0;		
DT	01-JUN-2002	(TREMBurel. 21,	
DT	01-JUN-2002	(TREMBurel. 21,	Last sequence update)
DT	01-JUN-2002	(TREMBurel. 21,	Last annotation update)
DE	SD02611p.		
GN	PNR.		

[illegible]

Mon Jan 6 08:17:56 2003

us-09-725-010-3.rspt

Page 6

Search completed: January 3, 2003, 19:50:35  
Job time : 49 secs







```
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KX Nuclear protein; Nitrate assimilation.
FT ZN_FING 665 689 GATA-TYPE.
SQ SEQUENCE 865 AA; 92407 MW; CC901BED187D8EEF CRC64;

Query Match 92.5%; Score 258; DB 1; Length 866;
Best Local Similarity 92.0%; Pred. No. 7e-25;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNPGQPLCNACGLFLKHLGVVRPLSLKTDVKKROR 50
    ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 665 CTNCFQTTPLRNPGQPLCNACGLFLKHLGVVRPLSLKTDVKKRNR 714

RESULT 5
AREA_ASPOP STANDARD; PRT; 866 AA.
ID AREA_ASPOP STANDARD; PRT; 866 AA.
AC OI3415;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN AREA.
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 4177;
RA Christensen T., Hynes M.J., Davis M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN POSITIVELY ACTING
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ002968; CAA05776.1; -
DR HSPD; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS00114; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KX Nuclear protein; Nitrate assimilation.
FT ZN_FING 664 688 GATA-TYPE.
SQ SEQUENCE 866 AA; 92882 MW; C09A03EB12E3FBB4 CRC64;

Query Match 92.5%; Score 258; DB 1; Length 866;
Best Local Similarity 92.0%; Pred. No. 7e-25;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNPGQPLCNACGLFLKHLGVVRPLSLKTDVKKROR 50
    ||||| :|||||:|||||:|||||:|||||:|||||:|||||
Db 664 CTNCFQTTPLRNPGQPLCNACGLFLKHLGVVRPLSLKTDVKKRNR 713

RESULT 6
AREA_EMENI STANDARD; PRT; 876 AA.
ID AREA_EMENI STANDARD; PRT; 876 AA.
AC P17429;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT -----
```

```
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90228331; PubMed=1970293;
RA Kudla B., Caddick M.X., Langdon T., Martinez-Rossi N.M.,
RA Bennett C.F., Sibley S., Davies R.W., Arst H.N. Jr.;
RT "The regulatory gene areA mediating nitrogen metabolism repression in
RT Aspergillus nidulans. Mutations affecting specificity of gene
RT activation alter a loop residue of a putative zinc finger.";
RL EMBO J. 9:1355-1364(1990).
RN [2]
RP REVISIONS.
RX MEDLINE=96123430; PubMed=8596437;
RA Langdon T., Seerins A., Ravagnani A., Caddick M.X., Arst H.N. Jr.;
RT "Mutational analysis reveals dispensability of the N-terminal region
RT of the Aspergillus transcription factor mediating nitrogen metabolite
RT repression.";
RL Mol. Microbiol. 17:877-888(1995).
RN [3]
RP STRUCTURE BY NMR OF 662-727.
RX MEDLINE=98202574; PubMed=9533883;
RA Starich M.R., Wikstroem M., Arst H.N. Jr., Clore G.M.,
RA Gronenborn A.M.;
RT "The solution structure of a fungal AREA protein-DNA complex: an
RT alternative binding mode for the basic carboxyl tail of GATA
RT factors.";
RL J. Mol. Biol. 277:605-620(1998).
CC -1- FUNCTION: TRANSCRIPTION ACTIVATOR THAT MEDIATES NITROGEN
CC METABOLITE REPRESSION IN A. NIDULANS. ACTIVATE THE TRANSCRIPTION
CC OF UAPA.
CC -----
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52491; CAA36731.1; -
DR PIR; S10017; S10017.
DR PDB; 4GAT; 28-JAN-98.
DR PDB; 5GAT; 28-JAN-98.
DR PDB; 6GAT; 28-JAN-98.
DR PDB; 7GAT; 28-JAN-98.
DR TRANSFAC; T02533; -.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS00114; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KX Nuclear protein; Nitrate assimilation; 3D-structure.
FT ZN_FING 673 697 GATA-TYPE.
FT DNA_BIND 721 742 H-T-H MOTIF (PROBABLE).
SQ SEQUENCE 876 AA; 94195 MW; 9ADC2273EE536F98 CRC64;

Query Match 92.5%; Score 258; DB 1; Length 876;
Best Local Similarity 92.0%; Pred. No. 7.1e-25;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNPGQPLCNACGLFLKHLGVVRPLSLKTDVKKROR 50
    ||||| :|||||:|||||:|||||:|||||:|||||:|||||
Db 673 CTNCFQTTPLRNPGQPLCNACGLFLKHLGVVRPLSLKTDVKKRNR 722
```

```

RESULT 7
AREA_ASPNG
ID AREA_ASPNG STANDARD; PRT; 882 AA.
AC 013412;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein area.
GN AREA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N402;
RA Maccabe A.P., Vanhanen S.A.S., Sollewijn Gelpke M.,
RA van de Vondervoort P., Arst H.N., Visser J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING
CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y11006; CAA71897.1; -
CC HSSP; P17429; 4GAT.
CC InterPro; IPR000679; Znf_GATA.
CC Pfam; PF00320; GATA; 1.
CC PRINTS; PR00619; GATAZNFINGER.
CC SMART; SM00401; Znf_GATA; 1.
CC PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
CC PROSITE; PS00114; GATA_ZN_FINGER_2; 1.
CC KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
CC Nuclear protein; Nitrate assimilation.
CC ZN_FING 694 718 GATA-TYPE.
CC FT NUCLEOTIDE SEQUENCE 971 AA; 103580 MW; 887DD882141C7453 CRC64;
CC
Query Match 92.5%; Score 258; DB 1; Length 882;
Best Local Similarity 92.0%; Pred. No. 7.1e-25;
Matches 46; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
CC
QY 1 CTNCGTKTTPLRNRRPQGPCLNACGLFLKLGHVVRPLSLKTDVIKKRQ 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 694 CTNCFQTTPLRNRRPQGPCLNACGLFLKLGHVVRPLSLKTDVIKKR 743
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
NIT2_NEUCR
ID NIT2_NEUCR STANDARD; PRT; 1036 AA.
AC P19212;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen catabolic enzyme regulatory protein.
GN NIT-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RA Fu Y.-H., Marzluf G.A.;
RA MEDLINE=90158568; PubMed=2137552;
RT "nit-2, the major nitrogen regulatory gene of Neurospora crassa,"
RT encodes a protein with a putative zinc finger DNA-binding domain."
RL Mol. Cell. Biol. 10:1056-1065(1990).
RN [2]
RP MUTAGENESIS.
RC MEDLINE=91186820; PubMed=2150539;
RA Fu Y.-H., Marzluf G.A.;
RA "Site-directed mutagenesis of the 'zinc finger' DNA-binding domain of
RT the nitrogen-regulatory protein NIT2 of Neurospora."
RL Mol. Microbiol. 4:1847-1852(1990).
CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. DURING CONDITIONS
CC OF NITROGEN LIMITATION IT TURNS ON THE EXPRESSION OF GENES FOR
CC ENZYMES WHICH ARE REQUIRED FOR THE USE OF A VARIETY OF SECONDARY
CC NITROGEN SOURCES, INCLUDING NITRATES, PURINES, AMINO ACIDS, AND
CC PROTEINS.
CC
```

```

CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- INDUCTION: BY LACK OF A PRIMARY NITROGEN SOURCE.
CC -|- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M33956; AAB03891.1; -.
DR PIR; A34755; A34755.
DR HSP; P17429; 4GAT.
DR TRANSFAC; T00627; -.
DR INTERPRO; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS0114; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein; Nitrate assimilation; Repeat.
FT DOMAIN 49 110 3 X APPROXIMATE REPEATS.
FT REPEAT 49 55 1.
FT REPEAT 87 92 1.
FT REPEAT 105 110 3.
FT ZN_FING 743 767 GATA-TYPE.
FT MUTAGEN 743 746 CTNC->STNG: ABOLISHES DNA-BINDING.
FT MUTAGEN 755 756 RR->GG: ABOLISHES DNA-BINDING.
FT MUTAGEN 765 766 NA->DV: ABOLISHES DNA-BINDING.
FT MUTAGEN 768 769 GL->DV: ABOLISHES DNA-BINDING.
FT MUTAGEN 789 790 KR->NS: ABOLISHES DNA-BINDING.
SQ SEQUENCE 1036 AA; 109310 MW; 5FE4992B1C223514 CRC64;

Query Match 91.8%; Score 256; DB 1; Length 1036;
Best Local Similarity 92.0%; Pred. No. 1.5e-24;
Matches 46; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNRRPQGLNACGLFLKLGHVVRPLSLKTDVIKKRQR 50
|||||1:|||||1:|||||1:|||||1:|||||1:|||||1:|||||1:|||||1:
DB 743 CTNCGTKTTPLRNRRPQGLNACGLFLKLGHVVRPLSLKTDVIKKRNR 792

RESULT 10
GAF1_SCHPO
ID GAF1_SCHPO STANDARD; PRT; 855 AA.
AC Q10280; O94482; O9USK9;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription factor gaf1 (Gaf-1).
GN GAF1 OR SPCC417.01C OR SPCC1902.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

```

```

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinhardt R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McComb W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 566-855 FROM N.A.
RX MEDLINE=98382525; PubMed=9714831;
RA Hoe K.-L., Won M.S., Chung K.-S., Park S.-K., Kim D.-U., Jang Y.-J.,
RA Yoo O.-J., Yoo H.-S.;
RT "Molecular cloning of gaf1, a Schizosaccharomyces pombe GATA factor,
RT which can function as a transcriptional activator.";
RL Gene 215:319-328(1998).
CC -|- FUNCTION: TRANSCRIPTIONAL ACTIVATOR.
CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -|- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL035076; CAA22647.1; -.
DR EMBL; AL049521; CAB40003.1; -.
DR EMBL; L31601; AAC35593.1; -.
DR HSP; P17429; 4GAT.
DR TRANSFAC; T02831; -.
DR INTERPRO; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS0114; GATA_ZN_FINGER_2; 1.
DR Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT ZN_FING 635 659 GATA-TYPE.
SQ SEQUENCE 855 AA; 91776 MW; 3D932F83D2DE6774 CRC64;

Query Match 86.7%; Score 242; DB 1; Length 855;
Best Local Similarity 84.0%; Pred. No. 7.2e-23;
Matches 42; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 CTNCGTKTTPLRNRRPQGLNACGLFLKLGHVVRPLSLKTDVIKKRQR 50
|||||1:|||||1:|||||1:|||||1:|||||1:|||||1:|||||1:|||||1:
DB 635 CTNCGTKTTPLRNRRPQGLNACGLFLKLGHVVRPLSLKTDVIKKRNR 684

RESULT 11
GAF1_YEAST
ID GAF1_YEAST STANDARD; PRT; 510 AA.
AC P43574;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcriptional regulatory protein GAT1.
GN GAT1 OR YFL021W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

```

```

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GRE88;
RX MEDLINE=96182087; PubMed=8622686;
RA Coffman J.A., Rai R., Cunningham T., Svetlov V., Cooper T.G.;
RT "Gatlp, a GATA family protein whose production is sensitive to
RT nitrogen catabolite repression, participates in transcriptional
RT activation of nitrogen-catabolic genes in Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 16:847-858(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yanazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -!- FUNCTION: POSITIVE REGULATOR OF MULTIPLE NITROGEN CATABOLIC GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U27344; AAB03516.1; -.
DR EMBL; D50617; BAA09217.1; -.
DR HSSP; P17429; 4GAT.
DR TRANSFAC; T02817; -.
DR SGD; S0001873; GAT1.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; DNA-binding; Zinc-finger; Nuclear protein.
FT ZN_FING 310 334
FT DOMAIN 151 158 POLY-ASP.
FT SEQUENCE 510 AA; 56327 MW; 62D805E42695F35F CRC64;
Query Match 85.3%; Score 238; DB 1; Length 510;
Best Local Similarity 82.0%; Pred. No. 1.4e-22;
Matches 41; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
YQ 1 CTNCGTKTTPLRNRNPGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRQ 50
Db 310 CSNCTSTTPLRKDPKGLPCNACGLFLKLGHVVRPLSLKTDVIKKRQ 359
RESULT 12
ASD4_NEUCR
ID ASD4_NEUCR STANDARD; PRT; 426 AA.
AC Q9HEV5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE GATA type zinc finger protein asd4 (Ascus development protein 4).
GN ASD-4.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=20455773; PubMed=10998244;
```

```

RA Feng B., Haas H., Marzluf G.A.;
RT "ASD4, a new GATA factor of Neurospora crassa, displays
RT sequence-specific DNA binding and functions in ascus and ascospore
RT development.";
RL Biochemistry 39:11065-11073(2000).
CC -!- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT FUNCTIONS IN SEXUAL
CC DEVELOPMENT; DISRUPTION OF ASD-4 GENE RESULTS IN AGENESIS OF ASCUS
CC AND ASCOSPORE WITH MACROSCOPICALLY NORMAL FRUITING BODY FORMATION.
CC THE GATA-TYPE ZINC FINGER DOMAIN BINDS TO DNA SEQUENCES FROM ITS
CC OWN PROMOTER REGION.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF319953; AAG45180.1; -.
DR HSSP; P17429; 4GAT.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 1.
DR PROSITE; PS50114; GATA_ZN_FINGER_2; 1.
KW Transcription regulation; DNA-binding; Zinc-finger; Nuclear protein;
KW Coiled coil.
FT ZN_FING 16 40 GATA-TYPE.
FT DOMAIN 182 292 COILED COIL (POTENTIAL).
FT SEQUENCE 426 AA; 46532 MW; 4B7DE193050E280E CRC64;
Query Match 75.3%; Score 210; DB 1; Length 426;
Best Local Similarity 76.0%; Pred. No. 4.1e-19;
Matches 38; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
YQ 1 CTNCGTKTTPLRNRNPGQPLCNACGLFLKLGHVVRPLSLKTDVIKKRQ 50
Db 16 CQNCATSTTPLRNDEMGGVLCNACGLFLKLGHRPRISLKTVDIKNR 65
RESULT 13
GLN3_YEAST
ID GLN3_YEAST STANDARD; PRT; 730 AA.
AC F18494;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nitrogen regulatory protein GLN3.
GN GLN3 OR YER040W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92049353; PubMed=1682800;
RA Minehart P.L., Magasanik B.;
RT "Sequence and expression of GLN3, a positive nitrogen regulatory gene
RT of Saccharomyces cerevisiae encoding a protein with a putative zinc
RT finger DNA-binding domain.";
RL Mol. Cell. Biol. 11:6216-6228(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunicker-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
```

RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,  
 RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,  
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: POSITIVE NITROGEN REGULATORY PROTEIN. REQUIRED FOR  
 CC THE ACTIVATION OF TRANSCRIPTION OF A NUMBER OF GENES (INCLUDING  
 CC THE ALLANTOIN PATHWAY GENES) IN RESPONSE TO THE REPLACEMENT OF  
 CC GLUTAMINE BY GLUTAMATE AS SOURCE OF NITROGEN. BINDS THE NITROGEN  
 CC UPSTREAM ACTIVATION SEQUENCE OF GLN1. THE GENE ENCODING GLUTAMINE  
 CC SYNTHETASE, URE2 MAY CATALYTICALLY INACTIVATE GLN3 IN RESPONSE TO  
 CC AN INCREASE IN THE INTRACELLULAR CONCENTRATION OF GLUTAMINE.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M35267; RAA34645.1; -;  
 DR EMBL; U18796; AAB64575.1; -;  
 DR PIR; S22280; S22280.  
 DR HSP; P17429; 4GAT.  
 DR TRANSFAC; T02818; -;  
 DR SGD; S0000842; GLN3.  
 DR InterPro; IPR000679; Znf\_GATA.  
 DR Pfam; PF00320; GATA; 1.  
 DR PRINTS; PR00619; GATAZNFINGER.  
 DR SMART; SM00401; Znf\_GATA; 1.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
 DR PROSITE; PS00114; GATA\_ZN\_FINGER\_2; 1.  
 KW Transcription regulation; Activator; DNA-binding; Zinc-finger;  
 KW Nuclear protein; Nitrate assimilation.  
 FT ZN\_FING 306 330 GATA-TYPE.  
 FT DOMAIN 351 361 ARG/LYS-RICH (BASIC).  
 FT CONFLICT 474 474 P -> G (IN REF. 1).  
 FT SEQUENCE 730 AA; 79382 MW; 3159E1844469942E CRC64;  
 SQ  
 Query Match 72.4%; Score 202; DB 1; Length 730;  
 Best Local Similarity 75.0%; Pred. No. 6.8e-18;  
 Matches 36; Conservative 4; Mismatches 8; Indels 0; Gaps 0;  
 OY 1 CTNCGTKTTPLRNRPQGLACGLFLKLGHVVRPLSLKTDVIKRR 48  
 DB 306 CFNCKTFTPLWRSPGNTLCACGLFQKLHGTMRPLSLKSDVIKRR 353  
 RESULT 14  
 ID GF3 YEAST STANDARD; PRT; 551 AA.  
 AC P42944;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE GF3 protein.  
 GN GF3 OR YJL110C OR J0806.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=96090136; PubMed=7483851;  
 RA Rasmussen S.W.;  
 RT "A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1  
 RT and CSD3 genes, a TCP-1-related gene, an open reading frame similar  
 RT to the DAL80 gene, and a tRNA(Arg).";  
 RL Yeast 11:873-883(1995).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=97103775; PubMed=8948101;  
 RA Cziapluch C., Kordes E., Pujol A., Jauniaux J.-C.;  
 RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X  
 RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,  
 RT SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, COT7, GZF3, two tRNA genes,  
 RT three remnant delta elements and a Ty4 transposon.";  
 RL Yeast 12:1471-1474(1996).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X85021; CAA59384.1; -;  
 DR EMBL; Z49385; CAA89405.1; -;  
 DR EMBL; X86353; CAA60126.1; -;  
 DR HSP; P17429; 4GAT.  
 DR TRANSFAC; T02827; -;  
 DR SGD; S0003646; GZF3.  
 DR InterPro; IPR000679; Znf\_GATA.  
 DR Pfam; PF00320; GATA; 1.  
 DR PRINTS; PR00619; GATAZNFINGER.  
 DR SMART; SM00401; Znf\_GATA; 1.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
 DR PROSITE; PS00114; GATA\_ZN\_FINGER\_2; 1.  
 KW Transcription regulation; DNA-binding; Zinc-finger; Nuclear protein.  
 FT ZN\_FING 131 155 GATA-TYPE.  
 FT SEQUENCE 551 AA; 60519 MW; 807243799BFAB704 CRC64;  
 SQ  
 Query Match 72.0%; Score 201; DB 1; Length 551;  
 Best Local Similarity 74.0%; Pred. No. 7e-18;  
 Matches 37; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
 OY 1 CPNCGTKTTPLRNRPQGLACGLFLKLGHVVRPLSLKTDVIKRR 50  
 DB 131 CKNCLTSTPLWRDRHGAMLCACGLFLKLGKPRISLKTVDIKSRNR 180  
 RESULT 15  
 ID PNR\_DROME STANDARD; PRT; 540 AA.  
 AC P52168;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE GATA-binding factor-A (Transcription factor GATA-A) (Pannier  
 DE protein).  
 GN PNR OR GATA-A.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=94139551; PubMed=7916677;  
 RA Winick J., Abel T., Leonard M.W., Michelson A.M.,  
 RA Chardon-Loriaux I., Holmgren R.A., Maniatis T., Engel J.D.;  
 RT "A GATA family transcription factor is expressed along the embryonic  
 RT dorsoventral axis in Drosophila melanogaster.";  
 RL Development 119:1055-1065(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94139568; PubMed=7916679;  
 RA Ramain P., Heitzler P., Haenlin M., Simpson P.;  
 RT "Pannier, a negative regulator of achaete and scute in Drosophila,  
 RT

Search completed: January 3, 2003, 19:48:37  
Job time : 15 secs